

Replacing Lecture with Web-Based Course Materials*

Richard Scheines,¹ Gaea Leinhardt,²
Joel Smith,³ and Kwangsu Cho,⁴

Abstract

In a series of 5 experiments in 2000 and 2001, several hundred students at two different universities with three different professors and six different teaching assistants took a semester long course on causal and statistical reasoning in either traditional lecture/recitation or online/recitation format. In this paper we compare the pre-post test gains of these students, we identify features of the online experience that were helpful and features that were not, and we identify student learning strategies that were effective and those that were not. Students who entirely replaced going to lecture with doing online modules did as well and usually better than those who went to lecture. Simple strategies like incorporating frequent interactive comprehension checks into the online material (something that is difficult to do in lecture) proved effective, but online students attended face-to-face recitations less often than lecture students and suffered because of it. Supporting the idea that small, interactive recitations are more effective than large, passive lectures, recitation attendance was three times as important as lecture attendance for predicting pre-test to post-test gains. For the online student, embracing the online environment as opposed to trying to convert it into a traditional print-based one was an important strategy, but simple diligence in attempting “voluntary” exercises was by far the most important factor in student success.

Acknowledgements

We thank David Danks, Mara Harrell and Dan Steel, all of whom spent hundreds of hours teaching, grading, and collecting the data that made these studies possible.

* This research was supported by the A.W. Mellon Foundation’s Cost-Effective Uses of Technology in Teaching program (<http://www.ceutt.org/>), the William and Flora Hewlett Foundation, and the James S. McDonnell Foundation.

¹ Dept. of Philosophy and Human-Computer Interaction Institute at Carnegie Mellon University.

² Learning Research and Development Center, and School of Education, University of Pittsburgh

³ Office of Technology for Education, and Chief Information Officer, Carnegie Mellon University.

⁴ Learning Research and Development Center, University of Pittsburgh

1. Introduction

Because courses given entirely or in part online have such obvious advantages with respect to student access and potential cost savings, their development and use has exploded over the last several years.⁵ Although we now know a little about online learning, e.g., how faculty and students respond subjectively to it and what strategies have proven desirable from both points of view (Hiltz et al., 2000; Kearsley, 2000; Sener, 2001; Wegner, et al., 1999; Clark, 1993; Reeves and Reeves, 1997; Song, Singelton, Hill, and Koh, 2004), we still know far too little about how online course delivery compares to traditional course delivery with respect to objective measures of student learning. Some studies have reported no significant difference in learning outcomes between delivery modes (Barry and Runyan, 1995; Carey, 2001; Caywood and Duckett, 2003; Cheng, Lehman, and Armstrong, 1991; Hilz, 1993; Russell 1999, Sankaran, Sankaran, and Bui, 2000), some have shown that online students fared worse (e.g., Brown and Liedholm, 2001; Wang and Newlin, 2000), some have found that online students fare better (Derouza and Fleming, 2003; Maki, Maki, Patterson, & Whittaker, 2000; Maki and Maki, 2002), but few have compared entire courses and still fewer have managed to overcome the many methodological obstacles to rigorous contrasts (Phipps, et al., 1999; Carey, 2001; IHEP, 1999).

Maki and Maki (2003, p. 198) point out that in comparisons that favor online delivery, “the design of the course (the instructional technology), and not the computerized delivery, produced the differences favoring the Web-based courses.” They also point out, however, that online courses can more readily enforce deadlines, thus encouraging more engagement with the material, they can offer student’s more immediate feedback, and they can make learning active, all features of the educational experience that we know improve learning outcomes.

In 5 experiments performed over 2000 and 2001, we compared a traditional lecture/recitation format to an online/recitation format, measuring learning outcomes and a variety of student behaviors that might explain differences in learning outcomes. We tried to remove all differences in the designs of the online and lecture versions of the course except those that are essential to the difference in the delivery modes, for example the immediate feedback and comprehension checks that are only available in online learning. In support of Maki and Maki (2003), we found that the immediate feedback

⁵ See, for example, the many efforts described or cited in (Bourne and Moore 2000)

and active learning clearly helped, but we also found that online students were less likely to attend recitation sections, which hurt. Overall, even controlling for pre-test and recitation attendance, we found that students in the online version of the course did slightly better than students in the lecture version of the course – independent of their lecturer, teaching assistant, gender, or any other feature we measured.

In the last of the experiments we discuss here, we recorded how many of the online modules each student chose to print out, and how many of the interactive exercises not available in the print-outs that they attempted. We found that those students who printed out modules did fewer interactive exercises and as a result fared worse on learning outcomes.

We do *not* want to argue that *interactive* face-to-face time between students and teachers should be replaced by the student-computer interaction – we believe no such thing. All of the students in our first year of experiments were encouraged to attend weekly face-to-face recitation sections, and all of the students in our second year were required to do so. The first question we are trying to address is the effect of replacing large lectures (e.g., over 50) with interactive, online courseware. In this paper, therefore, our priority is to address the simplest question about online courseware: can it replace large lectures without doing any harm to what the students objectively learn from the course. The second goal of this paper is to begin the process of identifying the features of online course environments that are pedagogically important, and the student strategies that are adaptive in the online setting and those that are not.

The paper is organized as follows. In the next section, we briefly describe the online course material. In section three we describe our experiments. In section four, we discuss the evidence for the claim that replacing lecture with online delivery did no harm and probably some good, and we discuss which features of the online environment helped and which seemed to hinder student outcomes. In section five we discuss the student strategies that were adaptive and those that were not, and in section six we discuss some of the many questions left unanswered and the future platform for educational research being developed by the Open Learning Initiative at Carnegie Mellon that will hopefully address them.

2. Online Courseware on Causal Reasoning

Although Galileo showed us how to use controlled experiments to do causal discovery more than 400 years ago, it wasn't until R.A. Fisher's (1935) famous work on experimental design that further headway was made on the statistics of causal discovery.

Done well before World War II, Fisher's work, like Galileo's, was confined to experimental settings in which treatment could be assigned. The entire topic of how causal claims can or cannot be discovered from data collected in non-experimental studies was largely written off as hopeless until about the mid 1950s with the work of Herbert Simon (1954) and the work of Hubert Blalock seven years later (Blalock, 1961). It wasn't until the mid 1980s, however, that artificial intelligence researchers, philosophers, statisticians and epidemiologists began to really make headway on providing a rigorous theory of causal discovery from non-experimental data.⁶ Convinced that at least the qualitative story behind causal discovery should be taught to introductory level students concurrent with or as a precursor to a basic course on statistical methods, and also convinced that such material could only be taught widely with the aid of interactive simulations and open ended virtual laboratories, a team at Carnegie Mellon and the University of California, San Diego⁷ teamed up to create enough online material for an entire semester's course in the basics of causal discovery. By the spring of 2004, over 2,600 students in over 70 courses at almost 30 different colleges or universities have taken all or part of our online course.

--- **Insert Figure 1 Here** ---

Causal and Statistical Reasoning (CSR)⁸ involves three components: 1) 17 lessons, or "concept modules" (e.g., see Figure 1), 2) a virtual laboratory for simulating social science experiments, the "Causality Lab"⁹ and 3) a bank of over 100 short cases : reports of "studies" by social, behavioral, or medical researchers taken from news service reports (e.g., see Figure 2).

--- **Insert Figure 2 Here** ---

Each of the concept modules contains approximately the same amount of material as a text-book chapter or one to two 90 minute lectures, but also includes many interactive

⁶ See, for example, Spirtes, Glymour and Scheines (2000), Pearl (2000), Glymour and Cooper (1999).

⁷ In addition to Scheines and Smith this includes Clark Glymour, at Carnegie Mellon and the Institute for Human-Machine Cognition (IHMC) in Pensacola, FL and David Danks, now at IHMC, Sandra Mitchell, now at the University of Pittsburgh, Willie Wheeler and Joe Ramsey, both at Carnegie Mellon.

⁸ CSR is available free at www.cmu.edu/oli.

⁹ The Causality Lab is available as a stand alone program: www.phil.cmu.edu/projects/causality-lab

simulations (e.g., see Figure 1), in some cases more extended exercises to be carried out in the Causality Lab, and frequent comprehension checks, i.e., two or three multiple choice questions with extensive feedback after approximately every page or so of text (e.g., the “Did I Get This?” link shown in Figure 1). At the end of each module is a required, graded online quiz.

The online material is intended to replace lectures, but not recitation. The online part of the course interactively and with infinite patience delivers the basic concepts needed to understand the subject, but human instructors possessing the subtle and flexible intelligence as of yet beyond computers lead discussion sections in which the basic concepts are integrated and then applied to real, often messy case studies.

3. The Experiments

The Treatments

In order to test the relative efficacy of delivering our material online, we created two versions of a full semester course, one to be delivered principally online and one principally by lecture. The two versions were as identical in all respects save delivery format as we could make them. In the online version of the course, students got the material from the online modules instead of lecture (they were required to complete one module each time a lecture was given on the same topic), and in fact were *not allowed* to go to lecture. At the end of each module is a required online mastery quiz, and students were required to exceed a 70% threshold on this quiz by a date just after the module was to be covered in recitation to get credit for having done the module. Their quiz grades and the dates of completion were available online to the TAs. Online students were encouraged to go to a weekly recitation in year 1, and were required to attend this recitation in year 2.

In the lecture version of the course, the class consisted of two lectures per week and a recitation section. For reading, the online modules were printed out (minus, of course, the interactive simulations and exercises) and distributed to the students. The lectures essentially followed the modules. Since the online version of the modules involved interactive simulations and exercises not included in the readings passed out to lecture students, extra assignments and traditional exercises approximating those given interactively online were given out to lecture students. As these exercises were voluntary in the online modules, they were also voluntary for the lecture students.

Both versions of the course included one interactive recitation section per week. Students were encouraged to bring up any questions they had with the material, and the TAs also handed out problem sets and case studies for the students to analyze and then discuss in the recitation. Since the mastery quizzes taken by online students were unavailable for lecture students, online students were dismissed 15 minutes early from the one hour recitation and lecture students were given a different but comparable version of the mastery quiz.

In three of the five experiments online and lecture students were assigned randomly to the same pool of recitations, but the results were indistinguishable to experiments in which online and lecture students were separated into recitation sections involving only students in their own treatment condition.

All students took identical paper and pencil pre-tests, midterms, and final exams, and they did so at the same time in the same room. The 18 item pre-test is a combination of six GRE analytic ability items (Big Book, Test 27) aimed exactly at the logic of social science methodology,¹⁰ four that tested arithmetic skills (percent, fractions, etc.), and eight that probed for background knowledge in statistics, experimental design, causal graphs, etc. Each midterm and the final was 80% multiple choice and 20% short essay, and in two experiments we graded them blind, which made no difference whatsoever.

We compared both delivery formats on a total of over 650 students, in five different semesters: 1) year 1: winter quarter in a Philosophy course on Critical Reasoning that satisfied a university wide requirement at UCSD (University of California, San Diego) 2) year 1: same course in the spring quarter at UCSD, 3) year 2: same course in the winter quarter at UCSD 4) year 2: same course in the spring quarter at UCSD, and 5) year 2:

¹⁰ For example: In an experiment, two hundred mice of a strain that is normally free of leukemia were given equal doses of radiation. Half the mice were then allowed to eat their usual foods without restraint, while the other half were given adequate but limited amounts of the same foods. Of the first group, fifty-five developed leukemia, of the second, only three.

The experiment above best supports which of the following conclusions?

- (A) Leukemia inexplicably strikes some individuals from strains of mice normally free of the disease.
- (B) The incidence of leukemia in mice of this strain which have been exposed to the experimental doses of radiation can be kept down by limiting their intake of food.
- (C) Experimental exposure to radiation has very little effect on the development of leukemia in any strain of mice.
- (D) Given unlimited access to food, a mouse eventually settles on a diet that is optimum for its health.
- (E) Allowing mice to eat their usual foods increases the likelihood that the mice will develop leukemia whether or not they have been exposed to radiation.

spring semester in a History and Philosophy of Science course on Scientific Reasoning that satisfied a university wide quantitative reasoning requirement at the University of Pittsburgh. The experiments involved three different lecturers, one who lectured both courses at UCSD in year 1, another who lectured both courses at UCSD in year 2, and a third who lectured at Pitt in year 2. The teaching assistants changed every semester.¹¹

Although we did not formally analyze the demographics of our students, they seemed representative of UCSD and Pitt with respect to race, gender, and ethnicity. The only exceptional characteristic seemed to arise from their relative lack of comfort with formal and analytic methods. In both cases the course satisfied a “quantitative or analytical reasoning” requirement, but was seen (we think incorrectly) as being less mathematically demanding than other courses that satisfied this requirement, e.g., a traditional Introduction to Statistics. Thus the students who participated were perhaps less comfortable with formal reasoning skills and computation than the mean in their cohorts – but in our view not substantially so.

Treatment Assignment

Allowing students to choose which delivery format they receive is desirable from the student’s point of view, but clearly invites a selection bias from our point of view, which is a disaster for causal inference. In fact most of the studies comparing online to lecture delivery that we are aware of did not randomize treatment assignment, even partially.¹² There are two simple ways to deal with treatment selection bias: randomly assign treatment or identify the potential source of the bias and then measure and statistically control for it. In year 1 we used a semi-randomized design, which employed both strategies (Figure 3).

---Insert Figure 3 Here ---

In year 1 we did not advertise the course as having an online delivery option. On the first day of class we administered a pre-test and informed students that they had the option to enter a lottery to take the course in the online format., which we explained. All students who wanted traditional lecture format (condition C) got it. We then took all the students who opted for the online delivery condition, ranked them by pre-test score, and then did a stratified random draw to give $2/3^{\text{rd}}$ of the students who wanted online delivery their choice: A) Online – wanted and got the online condition, and B) Control – wanted

¹¹ There was some overlap at UCSD in each year.

¹² See, for example, (Maki et al., 2000), and (Carey, 2001).

online but got lecture. Although this design leaves out one condition: students who wanted lecture but were assigned online delivery – we felt that such an assignment was unethical given how the course was advertised and given we did not yet know how the two groups would fare with respect to learning outcomes. We assured both groups that if there were any differences in the mean final course scores we would adjust the lower up by the difference in means.

In year 2, both at UCSD and at the University of Pittsburgh, students were again informed of the two options on the first day of class as well as how the previous year's groups had done, but the online option was advertised ahead of time, and all students were then given whichever treatment they chose.

4. Results

We present the results from these five experiments roughly chronologically, for several reasons. First, as with any experience that repeats, we learned things in early versions of the study, which we used to change later versions, and in several instances the lessons learned are worth recounting. Second, the scope and quality of the data collection effort improved over time. We had a richer set of measures to analyze in year 2, especially at Pitt. Finally, although presenting five studies sequentially may seem a little redundant, the fact that the results were approximately replicated over five slightly different versions of a course involving three different professors, six different teaching assistants, two different treatment assignment regimes, and two locations separated by over 2,500 miles convinced us far more than p-values that we were not seeing a statistical mirage. In what follows we slightly vary the format of our presentation of the results, mostly in response to the data available for the study reported on.

UCSD: Year 1

In the semi-randomized design used at UCSD in the winter and spring quarters of year 1 (Figure 3), two comparisons are in order: 1) the Online vs. Control comparison, and 2) the Control vs. Lecture comparison. Comparing Online vs. Control gives us the treatment effect among students who are disposed to do online courses, and comparing Control vs. Lecture gives us an estimate of the treatment selection bias, as these groups both received the same treatment (lecture delivery) but differed as to what delivery they chose.

--- Insert Figure 4 Here ---

Figure 4 displays the mean percents¹³ for each group on the pre-test, midterm and final exam and thus graphically summarizes the results for winter quarter, year 1. Pre-test means were statistically indistinguishable across groups, and although Online students outperformed Control and Lecture students, the differences were not significant at $\alpha = .05$, both in a simple difference of means test and in a regression in which we controlled for pre-test.¹⁴

Interestingly, although the Control and Lecture conditions showed literally no pre-test difference, Control students did consistently slightly outperform the Lecture condition by 2-4% – especially on the final exam ($p = .2$). We took this as suggestive evidence that there *was* a small selection bias of approximately 2-3% that our pre-test did not pick up. This is consistent with other studies comparing online vs. lecture treatment in which treatment was selected by the students and not assigned; see (Maki & Maki, 1997 and Maki, R. H., Maki, W.S., Patterson, M., & Whittaker, 2000), for example.

---- **Insert Figure 5 Here** ----

In the spring quarter, we repeated the experiment (Figure 5). Again, there was a small selection bias (2.7%), but unlike the winter quarter, in the spring quarter the Control condition consistently (albeit insignificant statistically) outperformed the Online condition. Upon examining the attendance records, a potential explanation emerged. Over the winter quarter, the lecture students attended an average of 85% of the recitations, but the online student attended an average of only 20%. In the spring, however, average recitation attendance among lecture students stayed at almost exactly 85%, but online students attended an average of fewer than 10% of recitations.

As a result of these experiments, we made two major modifications for year 2. First, because delivery choice and the pre-test were independent in year one, we allowed all students to choose their method of delivery, and second, we required recitation attendance of both online and lecture students. We again ran the experiment at UCSD in both winter and spring quarters of year 2, and also added a class in the spring semester of year 2 at the University of Pittsburgh.

¹³ All sample distributions were approximately normal.

¹⁴ Considering only items common to the pre-test and final exam, the online students did outperform the control group at $p = .015$.

UCSD: Year 2

The results in the winter quarter for year 2 at UCSD were quite similar to those in year 1, but in the spring quarter Online students showed a larger selection bias (3.3%) and larger performance advantage as well.

--- **Insert Table 1 Here** ---

Unfortunately, the connection between individuals and pre-test scores was corrupted in the year 2 winter data for UCSD, as was the attendance records, so only summary statistics are available. In the spring quarter, however, the Online students averaged 4.42% higher on the final exam than the Lecture students, after controlling for pre-test. Regressing Final exam score (in percent) on pre-test and a dummy variable to encode treatment condition (Online: 1= online, 0 = lecture), with standard errors in parentheses and p-values below gives the following results.

$$\begin{array}{l} \mathbf{Final\ (\%)} = 53.4 + 4.42\ Online + 0.315\ pre\text{-}test \\ \qquad\qquad\qquad (2.42) \qquad\qquad\qquad (0.087) \\ \qquad\qquad\qquad 0.073 \qquad\qquad\qquad **0.001 \end{array}$$

Maki and Maki (2002) found that higher multimedia comprehension skill predicted higher learning gains, and also interacted with web-based course format to predict learning gains. We did not find that cognitive ability (as measured by the pre-test) predicted higher learning *gains*, and we found no interaction between course delivery format and pre-test in predicting learning gains.

University of Pittsburgh: Spring semester Year 2

For several reasons, our best data come from the spring semester at the University of Pittsburgh. First, we were present to supervise data collection efforts. Second, and perhaps most importantly, we logged student behavior on a few important variables - how often they printed out modules to study, how often they attempted the voluntary comprehension checks inserted every page or two in the online modules, and how well they did on each post-module quiz.

As in the UCSD experiments performed in year 2, students were told the Online and Lecture options on the first day of class and then allowed to freely choose their treatment condition. At the University of Pittsburgh, 35 students chose Online and 50 chose Lecture. First, the difference in pre-test means between the Online and Lecture conditions was just over 1%, again statistically insignificant. Second, gender was independent of virtually every quantity we measured, including pre-test, treatment preference, and exam performance, thus it can be left out of our statistical analysis of the causes of exam performance. Third, dropout, which averaged around 10-15% across our experiments, was nearly independent of treatment condition and thus had little or no effect on any of estimates of treatment effect.

After controlling for pre-test and recitation attendance, online students averaged 5.3% higher on the final exam. Regressing Final exam score (in percent) on pre-test, the percentage of recitations attended, and a dummy variable for treatment condition gives the following results:

$$\begin{array}{l} \mathbf{Final} = 42.1 + 0.220 \text{ pre-test} + 0.233 \text{ rec} + 5.26 \text{ Online} \\ \qquad \qquad \qquad (.092) \qquad \qquad \qquad (.069) \qquad \qquad \qquad (2.73) \\ \qquad \qquad \qquad *.021 \qquad \qquad \qquad **.001 \qquad \qquad \qquad .059 \end{array}$$

Consistent with the UCSD experiment in spring of year 2, Online is significant at .1 but not at .05. It also shows that, as we had suspected from the UCSD experiments, recitation attendance strongly predicts final exam performance. Even controlling for pre-test and course delivery format, the expectation of Final exam score rises almost a quarter point (.233) for every extra percent of recitation attendance. Since there were only 13 recitations, each accounting for almost 8% of total recitation attendance, each extra recitation session attended increases the expectation for the Final exam by almost 2%.

To get a further handle on the importance of recitation attendance, we compared the relative importance of recitation vs. lecture attendance among only Lecture students in

the Pitt experiment. These students were supposed to attend lecture twice a week and recitation once, but attendance at recitation was over four times more predictive than attendance at lecture in a regression with Final as the dependent variable:

$$\begin{array}{rcl} \mathbf{Final}^{15} = & .317 \text{ rec\%} + & 0.078 \text{ lec\%} \\ & (.117) & (.101) \\ & ** .010 & .448 \end{array}$$

We take this as evidence, found by many others, that students learn more from small sessions in which they are active and engaged as opposed to large lectures in which they are for the most part passive.

Although the percent of recitations attended among online students rose from an embarrassing average of 8.6% in the spring of 2000 at UCSD to an average of 71% in the spring of 2001 at Pitt, it still trailed average recitation attendance among Lecture students (81%) by 10% ($p = .05$). We hypothesize that this discrepancy is a result of the greater aversion among those students who chose online to attend scheduled educational gatherings. It might, however, be the result of reduced weekly contact, or the greater independence required of online students. We do not yet know. If being in the online condition caused students to attend fewer recitations, then that probably has an adverse indirect effect on performance.

Path Analysis

Since there might well be two mechanisms through which treatment effects learning outcome, one direct and the other indirect, we used path analysis (Wright, 1921; Bollen, 1989) to estimate the strength of each mechanism. Table 2 shows the sample correlations among the four variables, with an “*” attached to correlations significant at .05:

Pre: pre-test%,

Online: (1 = yes, 0 = no),

Rec: % recitations attended

Final: final exam %

¹⁵ Only among students who chose the Lecture condition.

---- Insert Table 2 Here ---

The path model we used to estimate the relations among these variables is shown in Figure 6, along with the path coefficients, estimated not from the correlations but from the raw data to connect easily to regression results above. The path coefficients on the edges going into Final are almost identical to the regression estimates shown above.

--- Insert Figure 6 Here ---

The path model as a whole contains two important pieces of information. First, the fact that there are no edges connecting Pre to either Online or Rec is important, as it signifies that ability as measured by the pre-test has no influence on treatment selection, and no influence on whether a student attends recitation. Second, there are two paths from treatment (Online) to Final. The direct path indicates that, controlling for pre-test and recitation attendance, online students tend to average 5.26% higher on the Final than Lecture students. The indirect path: Online → Rec → Final, however, indicates first that Online students attend 10.2% fewer recitations on average, but that each extra percent of recitation attendance increases a student's average final exam score by .23%, meaning the indirect effect of Online on Final through Rec is to reduce Final exam scores by an average of 2.38 percent. Thus, if the path model above is correctly specified, the total effect of Online on Final exam score is $5.3 - 2.4 = 2.9$, or about a 1/3 of a grade.

The standard approach to estimating the strength of the relationships between variables like these, is to first specify a statistical model, and then calculate p-values relevant to the existence of particular relationships. This sort of statistical inference, however, is *conditional* on the model specification, a fact that is appreciated in theory but widely ignored in practice. Put another way, coefficient estimates and standard errors will vary considerably with the model specification, so unless one has high confidence in the model specification, the statistics are illusory. With a p-value of .96, the path model shown in Figure 6 fits the data extremely well,¹⁶ and in an exhaustive search of all possible alternative path models consistent with the time order among the variables in this model,¹⁷ no alternative fit as well.

¹⁶ In path analysis, higher p-values mean better overall fit. See (Bollen, 1989).

¹⁷ Pre-test was prior to treatment selection, which was prior to recitation attendance, which was prior to final exam.

Path models are limited in that they do not allow for the possibility of unmeasured confounders. In this case, the significant negative correlation between Online and recitation attendance might be due to an unmeasured confounder and not the result of a direct cause, but we modeled it as a direct cause because if anything this specification *weakened* the case for Online being the better treatment condition.

5. The Good Online Student

Up to this point we have compared the learning outcomes of online vs. lecture students. In this section we begin the process of analyzing the sorts of student behaviors that support or restrict objective learning in the online setting.

As with face-to-face instruction in colleges and universities there is a presumed set of student behaviors and an enacted set of behaviors. The presumption is that students will want to maximize their learning outcomes in a given course and toward this end attend classes, do the suggested readings at a fairly steady pace, do the homework as assigned, and study for tests and quizzes in a way that integrates new pieces of information together in a coherent and flexible fashion. In other words the student is expected to become engaged with both the process and substance of a course. Becoming engaged is somewhat more ill defined in the online setting. One might hypothesize that the skills of studentship are the same online as they are in a face-to-face setting. It might be the case, however, that in the online setting students adopt a more passive ‘just follow the directions’ stance, or it might be that online course work requires a more engaged and active student - one that moves around flexibly in the virtual world as opposed to linearly in a textbook world. The good online student might want to become engaged, but it might not always clear how they are to engage with an online course.

One mechanism we investigated involves extracting the material from the screen and placing it on paper. The paper version can be marked up, shuffled around, carried, and studied in a variety of environments. While analogous activities can be carried out on the computer they are more effortful and often less satisfactory – scribbling in the margins and drawing small diagrams does not require opening new windows or highlighting and dragging. McIsaac and Gunawardena (1993) suggest that print is a critical support for distance learners in current online learning systems. We also know that a good indication of engagement is that students actually do the embedded problems of the course material (Pressley & Ghatala, 1990) and do not simply flounder by clicking answers until they find the right ones.

A good student in this course would need to find a way to access and notate the material, study the examples carefully, take all of the embedded questions and note them, study the materials sufficiently carefully to pass the quizzes at the end of each module. One way to access and notate the materials is to print them out. However, when the modules are printed out the embedded questions and interactive material disappear. Thus the student must read/study the print-out off line and take the embedded questions and run the simulations or study the materials online separately. We began to study some of these issues in the spring of 2001 by recording more about student behavior than just attendance, pre-test, post-test, etc.

Population

The study was conducted in the first half of year 2 and involved two groups of online students, one taking the course in the winter quarter at UCSD and the other at the University of Pittsburgh in the spring. Out of the 75 students who decided to take the course online and who stayed in the course for the entire semester 68 records were obtained, 52 of which were complete and used here.

Measures

Pretest (Pre). A combination of GRE items that tested the sort of analytic ability germane to the material as well as items similar to those on the midterm and final exams. The score is the percent correct.

Printout usage (Print). As a feature of the courseware each module has available a “print” feature/link. If a student clicks on this button then this links the student to a ‘printable’ page made up of the entire module and its headings. Therefore, whenever a student made use of this feature a record of the behavior was available to us. The ‘printout’ measure consists of a ratio of the total number of clicks to this button divided by the total number of modules accessed by the student. Printing out the module is a mixed signal, as it indicates a level of engagement, but perhaps a resistance to using the modules online, as they were intended. Obviously printing in and of itself does nothing to the acquiring of knowledge.

Voluntary Questions Attempted (VolQs). As we described above each module contains a set of embedded comprehension checks. These questions probe the students on material introduced in approximately the previous ten to fifteen minutes. Sometimes the questions follow an active simulation. Ideally a student would run all of the

simulations in the module and would answer all of the embedded questions. However, a student might choose to skip the questions, not do the simulation, answer the questions by scanning all of the possible answers, etc. More important even than these problems is the fact that if a student is working from a printout version then it requires extra effort to do any of the embedded questions. The measure of voluntary questions attempted is the ratio of the number of embedded questions actually attempted divided by the total number of embedded questions that could have been attempted.

Quiz score (Quiz). Each module ends with a quiz. The students take the required quiz online and they received a percentage correct score. The quiz score contributed to their course grade. We summed the percentage correct divided by the number of quizzes taken over the entire course to construct a measure of average total quiz score.

Final exam score (Final). This is the student's percentage correct on the final exam.

Path Analytic Models

As above, we used path analytic models to study the causal relationships among these variables. Correlations, means, and standard deviations for these measures are given in Table 3.

--- **Insert Table 3 Here** ---

Unlike the path model in Figure 6, where we had good scientific reason to prefer a model specification we could then test and compare against alternatives, in the case here, even after assuming the relationships are approximately linear,¹⁸ we do not have sufficient domain knowledge to specify a unique path model among the five variables above. A variety of approaches exist to handle specification uncertainty. One can articulate a list of plausible models, assign a degree of belief to each, and then model average to compute the appropriate estimates and confidence intervals. This is only feasible for a small set of alternative specifications over which one has coherent degrees of belief, again a luxury we do not have here. One can also search among the model specifications considered equally plausible, and report on features shared by those models which best fit the data. We take this approach.

The variables above were measured in the same order in which we list their abbreviations, so we searched the 2^{10} path analytic models consistent with this time order,

¹⁸All variables reasonably approximate a normal, or truncated normal distribution.

using the PC and GES algorithms described in Spirtes, Glymour, & Scheines (2000) and implemented in TETRAD 4.¹⁹ The model in Figure 7 is the clear favorite.²⁰ With a p-value of .42, which is a measure of goodness-of-fit in path models and thus better when higher, this model fits the data quite well.

--- Insert Figure 7 Here ---

The most important coefficients, those expressing the direct influences on *Final*, are given in Table 4.

--- Insert Table 4 Here ---

Coefficients representing the relationships between the same predictors, but with *quiz* as the dependent variable, are given in Table 5.

---- Insert Table 5 Here ----

Other models that do well in the search are mostly variants of Figure 7 that simplify the model by removing edges that are marginally significant (represented as dashed lines in Figure 7). None of the estimates on the edges that remain change dramatically, which inspires confidence. We list several of the top models in Table 6 by indicating which edge in was dropped from the model in Figure 7, and the corresponding change in model fit statistics.

--- Insert Table 6 Here ---

It is the set of features that are shared by the top models that warrant confidence. Several are worth noting. First, attempting the embedded optional questions (*Volqs*) raised a student's average quiz scores dramatically. The coefficient representing this relationship is estimated at from .7 to .8 for all the top models. This means that the percentage of optional questions skipped accounts for approximately 2/3 of the variance in quiz score, even controlling for pre-test.²¹

¹⁹ TETRAD 4 is freely available: www.phil.cmu.edu/projects/tetrad

²⁰ All variables were standardized to have mean 0 and standard deviation of 1 before analysis.

²¹ One might reasonably ask whether all or some of the arrows in these models, which represent direct causal influence, could be replaced by latent common causes. In the case of the edge from *Volqs* to *Quiz*,

Second, the effect of the pre-test on the final exam is quite stable in all the top models. Estimates are all significant, and range from approximately .25 to .35. Clearly, pre-course knowledge and ability predicts performance on the final exam. Third, no top model postulates a direct connection from quiz to final. In each case, the association between *Quiz* and *Final* (.399, $p < .05$), is mediated mostly by *Volqs*, in others by *Volqs* and *Pre*, and in a few by *Volqs*, *Pre*, and *Print*. Fourth, estimates of the effect of *Volqs* on *Final* range from .35 to .44, and are significant in each case.

What, from the perspective of trying to characterize the good online student, do these results mean? First, the good student takes advantage of the frequent voluntary comprehension checks (with feedback) embedded every page or two in the online modules. Printing the modules is in conflict with engaging the interactive exercises, which means it has at least an indirectly negative effect on both quiz and final exam performance. Its direct effect, however, is harder to gauge. The literature suggests, and our data supports, that good students (as indicated by pretest score) *do* print out textual material originally available online. Our data, however, support a more complicated story. Even after controlling for *Pre* and *Volqs*, the effect of *Print* on *Quiz* and *Final* is negative, although not significantly so ($p = .15$ and $.12$ respectively). We cautiously hypothesize the following mechanism. Students who choose to print often are engaged and enthusiastic, yet are probably taking a different strategy for studying for the quizzes and exams. They most likely make notes on their printouts and consult these notes and the printed text while studying. They may also have a different pattern of studying, one that is more like that of cramming the material all at once which would account for the weak but consistent negative link between it and final test. The students who did not print the modules probably studied by revisiting the interactive questions with feedback, re-doing the simulations, a behavior which we unfortunately did not record. There are two plausible pathways from printing to performance that do not go through *Volqs* or *Quiz*: one through note taking and highlighting, the other through revisiting the interactive exercises. Printing encourages the first and discourages the second, thus the overall effect for this version of the online material was mildly negative. Revisiting the

there are grounds for denying this possibility. All of the top models that do not include the *Pre* \rightarrow *Quiz* and the *Print* \rightarrow *Quiz* edges entail that the association between *Print* and *Quiz* vanishes conditional on *Volqs*. Along with this statistical claim, and the supposition that *Print* is prior to *Volqs*, *Print* is an instrumental variable (Bowden and Turkington, 1984; Scheines et al., 2001) for the effect of *Volqs* on *quiz*, making the path analytic coefficient of *Volqs* on *quiz* an unbiased estimate of the causal effect, even if *Volqs* and *quiz* are confounded (Pearl, 2000).

interactive exercises is the instructor selected emphasis, but that may not be the student's choice.

6. Conclusions

After five different experiments involving over six hundred students, three different lecturers, six different recitation instructors, and two different locations, we are convinced that replacing lecture with online material did no harm and probably some good. Given that our online material is far from optimal, and that we have just begun to systematically collect data on student behavior and performance that will help us improve it, we believe that we will soon be able to show a significant gain from using interactive online material in place of lecture, precisely because of the opportunities online learning affords for more immediate feedback, more active learning, and more personalized opportunity to engage the material when and how the student wishes.

Although students who gave up attending lecture to learn online did as well, face-to-face contact at a weekly recitation section was clearly essential. For every recitation attended, a student can expect an extra 2% on the final exam. Among those who received traditional lecture/recitation delivery format, recitation attendance was four times more predictive of final exam score than lecture attendance. In our experience, however, online students were less likely to attend recitation. Even when it was the only face-to-face contact online students engaged in for a whole week but one of three face-to-face contacts for lecture students, online students still attended an average of 71% of recitations compared with 81% for lecture students. We do not yet know how to explain this difference.

The online environment is different than the traditional one, and it is not immediately clear from the student's perspective how best to learn within it. Unfortunately, most students put effort into only those activities they consider causally efficacious for their grade. In our case, students were asked to work through the modules, do the frequent comprehension checks as well as the simulations/animations/labs within the modules, and come to recitation to discuss the case studies. As it turned out, taking full advantage of the frequent but voluntary comprehension checks embedded in our modules was crucial, but the students didn't realize it, as the average student attempted only 50% of them. We do not yet know how important voluntary use of the simulations/animations/labs proved to be, but we suspect their use was highly correlated with the use of the comprehension checks. Some researchers, e.g., Pane, Corbett, and John, 1996, have found that

presenting animations or dynamic simulations has little effect on objective learning, however, so we are unsure of the effect we will find.

A service we provided but did not anticipate making much of a difference was the ability to print out the modules stripped of comprehension checks and all interactive material. We believe that good students saw this as an opportunity to further engage the material, but as it turns out printing came at a price: students who printed out the modules frequently tended not to go back and do the comprehension checks and/or the simulations/animations/labs, and our data seem to indicate that their performance on quizzes and final exams suffered accordingly.

We need to build online environments that support students, not only with content and interactivity, but also in how they are using the environment itself. In future versions of our course, for example, the computer will inform the student if he or she is exhibiting certain behaviors we are convinced are not adaptive. If, for example, a student is printing out the modules but not doing the voluntary comprehension checks and also not doing well on the post-module quizzes, then he or she should be informed that such a strategy is counterproductive, not in terms of printing, but in terms of forgoing the comprehension checks and interactive material.

Finally, because online technology makes it possible to automatically collect a lot of data, it presents as large an opportunity to do research on effective teaching and learning as it does to increase access and to reduce costs. We should be able to answer questions like: in an online course, where are students really spending their time? Where are they really having trouble? What sorts of activities are helpful? When and for what sorts of activities do they need face-to-face time, and when will online learning suffice?

Seeing the potential for using online courseware to learn about learning, we recently joined a larger effort at Carnegie Mellon, the Open Learning Initiative (OLI),²² which is building infrastructure with the dual purpose of delivering high quality online educational material and of supporting iterative improvement by including tools with which to systematically study online learning. The OLI's course delivery system will log every student move in a web-course, and, if desired, trace their path through a problem solving exercise in a virtual lab. The software is now being used to trace student moves in a virtual lab for aqueous chemistry, for proof construction in formal logic, for exploratory data analysis in introductory statistics, for supply and demand experiments in economics, and for simulated social science experiments. The results we report here are the tip of the

²² See <http://www.cmu.edu/oli/>.

iceberg of what is now becoming technologically feasible through collaborative efforts like OLI.

7. References

- Barry, M., and Runyan, G. (1995). A review of distance learning studies in the U.S. military. *The American Journal of Distance Education*, 9, 3, 37-47.
- Bates, A., and Poole, G. (2003). *Effective Teaching with Technology in Higher Education*, Jossey-Bass, San Francisco, CA.
- Blalock, H. (1961). *Causal Inferences in Nonexperimental Research*. University of North Carolina Press, Chapel Hill, NC.
- Bollen, K. (1989). *Structural Equations with Latent Variables*. Wiley, New York.
- Bourne, J. and Moore, J. (2001). *Online Education: Proceedings of the 2000 Summer Workshop on Asynchronous Learning Networks, Volumes I and II*, Sloan Center for Online Education.
- Bowden, R. and Turkington, D. (1984). *Instrumental variables*. Cambridge University Press, NY
- Brown, A.L., & Campione, J.C.(1994) Guided discovery in a community of learners. in *Classroom Lessons: Integrating Classroom Practices*, K. McGilly, ed. Cambridge, MA., MIT Press
- Brown, B., and Liedholm, C. (2002). Can Web-courses replace the classroom in Principles of Microeconomics. *American Economic Review*, May.
- Carey, J. (2001). Effective student outcomes: A comparison of online and face-to-face Delivery Modes. *Distance Education Online Symposium*
- Caywood, K., ,and Duckett, J. (2003). Online vs. on-campus learning in teacher education. *Teacher Education and Special Education*, 26, 2, 98-105.
- Cheng, H., Lehman, J., and Armstrong, P. (1991). Comparison of performance and attitude in traditional and computer conferencing classes, *The American Journal of Distance Education*, 5, 3, 51-64.
- Cho, K., Schunn, C. D., & Lesgold, A. (2002). Comprehension monitoring and repairing in distance collaboration. In the *Proceedings of the 24th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.
- Cho, K., & Schunn, C. (2003). Seven cognitive factors in networked collaboration. In the *Proceedings of the 25th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.
- Clark, T. (1993). Attitudes of higher education faculty toward distance education: a national survey. in *American Journal of Distance Education*. 10, 2, 4-36.
- Collins, A. (1991). Cognitive apprenticeship and instructional technology. In L. Idol and B. F. Jones (Eds.). *Educational values and cognitive instruction: Implications for reform* (pp. 121-138). Hillsdale, NJ: Erlbaum.

Derouza, E. and Fleming, M. (2003). A comparison of in-class quizzes vs. online quizzes on student exam performance. *Journal of Computing in Higher Education*, 14, 2, 121-134.

Educational Testing Service (1995). *GRE: Practicing to Take the General Test : Big Book*. Educational Testing Service Fisher, R.A. (1935). *The Design of Experiments*. Hafner Publishing Co., New York

Glymour, C., and Cooper, G. (1999). *Computation, Causation, and Discovery*. AAAI Press and MIT Press.

Greeno, J.G. (1998). Where is teaching. *Issues in Education*, V. 4, No. 1, 125-132.

Hiltz, S. (1993). Correlates of learning in a virtual classroom. *International Journal of Man-Machine Studies*, 39, 71-98

Hiltz, S.R., Benbunan-Fich, R., Coppola, N., Rotter, N. & Turoff, M. (2000). Measuring the importance of collaborative learning for the effectiveness of ALN: A multi-measure, multi-method approach. *Journal of Asynchronous Learning Networks*, 4, 2 (www.aln.org/alnweb/journal/jaln-vol4issue2-3.htm)

Institute for Higher Education Policy. (1999). What's the difference? A review of contemporary research on the effectiveness of distance learning in higher education. Available at <http://www.ihep.com/Pubs/PDF/Difference.pdf>

Kearsley, G. (2000). *Online Education: Learning and Teaching in Cyberspace*, Wadsworth, Belmont, CA.

Koedinger, K. R., & Anderson, J. R. (1998). Illustrating principled design: The early evolution of a cognitive tutor for algebra symbolization. *Interactive Learning Environments*, 5, 161-180.

Leinhardt, G. (1980). Modeling and measuring educational treatment in evaluation. *Review of Educational Research*, 50(3), 393-420.

Maki, W.S., & Maki, R. H. (1997). Learning without lectures: A case study, *IEEE Computer*, 30, 107-108

Maki, W.S and Maki, R. H. (2000). Evaluation of a Web-based introductory psychology course: II. Contingency management to increase use of online study aids. *Behavior Research Methods. Instruments & Computers*, 32. 240-245.

Maki, R. H., Maki, W.S., Patterson, M., & Whittaker, P.D. (2000). Evaluation of a Web-based introductory psychology course: I. Learning and satisfaction in online vs. lecture courses. *Behavior Research Methods. Instruments & Computers*, 32. 230-239.

Maki, W.S., & Maki, R. H. (2002). Multimedia comprehension skill predicts differential outcomes of Web-based and lecture courses. *Journal of Experimental Psychology: Applied*, 8, 85-98.

Maki, R., and Maki, W. (2003). Prediction of learning and satisfaction in Web-based and lecture courses. *Journal of Educational Computing Research*, 28, 3, 197-219.

Mayer, R. E. & Anderson, R. B. (1991) Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology* 83, 484-490.

- Mayton, G. B. (1990). The effects of the animation of visuals on the learning of dynamic processes through microcomputer-based instruction. Ohio State University: *PhD thesis*.
- McIsaac, M. S., & Gunawardena, C. N. (1993). Distance education. D. H. Jonassen (ed.), *Handbook of research; Educational communications and technology*. New York: Simon & Schuster Macmillan.
- Nardi, B. A. (1996). Studying context: A comparison of activity theory, situated action models, and distributed cognition. In B. A. Nardi (Ed.). *Context and consciousness: Activity theory and human-computer interaction*. Cambridge, MA: MIT Press.
- Pane, John F., Corbett, Albert T., John, Bonnie E., (1996). Assessing dynamics in computer-based instruction”, School of Computer Science, Carnegie Mellon, Pittsburgh, PA, http://www-2.cs.cmu.edu/%7Eacse/chi96_electronic/
- Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge University Press. Cambridge, UK.
- Phipps, Merisotis, and O'brien (1999). “What's the difference?: A review of contemporary research on the effectiveness of distance learning in higher education. Washington, DC: *The Institute for Higher Education*. (<http://www.ittheory.com/difference.pdf>)
- Pressley, M., & Ghatala, E. S. (1990). Self-regulated learning: Monitoring learning from text. *Educational Psychologist*, 25(1), 19-33.
- Reeves, T. and Reeves, P. (1997). Effective dimensions of interactive learning on the World Wide Web,” in *Web-based Instruction*, Badrul H. Khan (ed.), Educational Technology Publications, Englewood Cliffs, NJ.
- Russell, T. (1999). *The No Significant Difference Phenomenon*. Office of Instructional Telecommunications, North Carolina State University.
- Sankaran, Sankaran, and Bui (2000). Effect of student attitude to course on learning performance: an empirical study in Web vs. lecture instrucion. *Journal of Instructional Technology*, 27, 66-73.
- Sener, J. (2001). Bringing ALN into the mainstream: NVCC case studies, in *Online Education: Proceedings of the 2000 Summer Workshop on Asynchronous Learning Networks, Volume II*, Bourne, J. and Moore, J. (eds.), Sloan Center for Online Education.
- Simon, H. (1953). Causal ordering and identifiability. *Studies in Econometric Methods*. Hood and Koopmans (eds). 49-74. Wiley, NY.
- Song, L., Singelton, E., Hill, J., and Koh, M (2004). Improving online learning: Student perceptions of usefui and challenging characteristics. *The Internet and Higher Education*, 7, 1, 59-70.
- Spirtes, P., Glymour, C., Scheines R., (2000). *Causation, Prediction and Search*, 2nd Edition, MIT Press, Cambridge, MA.
- VanLehn, K., Siler, S., Murray, C, Yamauchi, T. & Baggett, W. B. (in press). Human tutoring: Why do only some events cause learning? *Cognition and Instruction*.

VanLehn, K., Freedman, R., Jordan, P., Murray, C., Osan, R., Ringenberg, M., Rose, C., Schulze, K., Shelby, R., Treacy D., Weinstein, A., and Wintersgill, M. (2000). Fading and deepening: The next steps for Andes and other model-tracing tutors. in C. Frasson (ed.), *Proceedings of ITS 2000*. Berlin: Springer-Verlag.

Wang, Y., and Newlin, M.(2000) Characteristics of students who enroll and succeed in psychology Web-based classes. *Journal of Educational Psychology*, 92, 137-143.

Wegner, S, Holloway, K., and Garton, E. (1999). The effects of internet-based instruction on student learning. *Journal of Asynchronous Learning Networks*, 3, 2.

Wright, S. (1934). The method of path coefficients. *Ann. Math. Stat.* 5, 161-215.

Direct reprint requests to:

Dr. Richard Scheines

Department of Philosophy

Carnegie Mellon University

Pittsburgh, PA 15213

email: scheines@cmu.edu

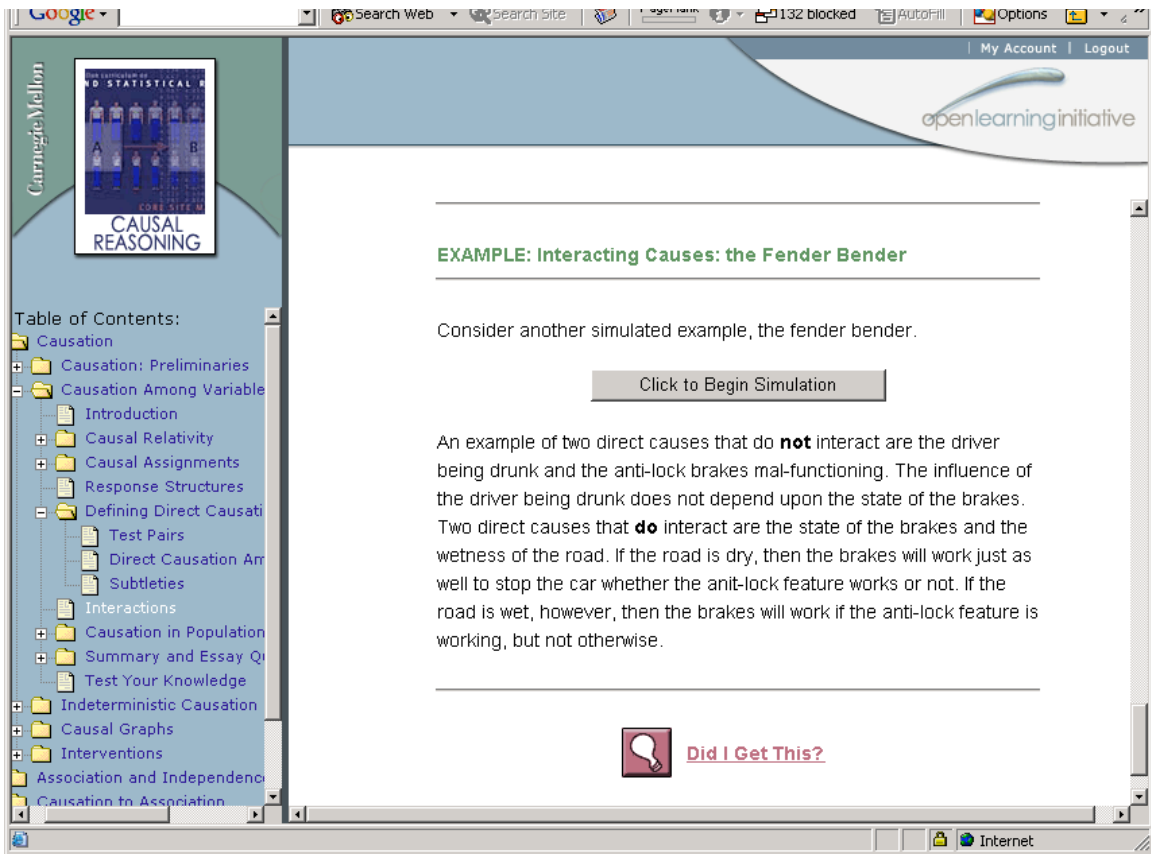


Figure 1: CSR Module Screen Shot

80-321 Causation and Social Policy - Microsoft Internet Explorer

Curran/Mellon

CAUSAL REASONING

- Study Suggests Breast Cancer Is Linked to Use of Antibiotics
- Exercise 'prevents breast cancer'
- Breast Cancer Link to Working Nightshifts
- Breast Cancer Care Worse for Poor
- Breast Density Study
- Breast-Feeding Reduces Teen Obesity
- Study: Cell phones not linked to cancer
- Pennsylvania cell-phone law put on hold
- The Urge to Punish Cheats: It Isn't Merely Vengeance
- Not Entirely Clear: In Virginia, Water Chlorination Is Blamed for
- Ask the Experts: Health benefits of circumcision?
- Colombian cities experiment with gun ban
- Computer games make children anti-social
- Cot death babies 'dreaming of womb'
- Budding Couch Potatoes
- Link Found Between Behavioral Problems and Time in Child Ca
- Why Are So Many Women Depressed?
- Vaginal Birth Tougher for Women with Diabetes
- Exploring the Diabetes-Depression Link
- Sugar and Spice and Everything Nice ...
- Early education helps at risk youth
- Exercise 'could halt mental decline'
- Accident risk for fat drivers
- Fetal Exposure to Alcohol Linked to Mental Illness
- Report Finds 2000 Florida Vote Rife with Disparities
- Flu shot lowers risk for second heart attack, study shows
- Forgiveness is good for the heart and soul
- Friday 13th - unlucky for some
- Winners and Losers: THE BENEFITS OF GLOBALIZATION ARE S
- Weighing the Grandma Factor; In Some Societies, It's a Matter
- Short Changed: Why do tall people make more money?
- Grave reveals secret of long life
- Mutation protects against HIV/AIDS

openlearninginitiative

My Account | Logout

So it may come as no surprise that Johns Hopkins University researchers are pointing the finger squarely at TV as the primary cause of obesity in American youngsters, especially in girls and minority children.

In a study published March 25 in the Journal of the **American Medical Association**, they report that teenage girls and black and Hispanic kids of both genders watch more TV than their white, male peers, and they're also heavier.

"There's no question that these kids are not out burning calories," says obesity researcher Ross Andersen, lead author of the study.

He and his colleagues compared the TV-watching habits of 4,063 children between 8- and 16-years-old with their physical activity levels and body-fat measurements.

Regardless of race or ethnicity, those who watched four or more hours of TV each day had a greater proportion of body fat than kids who watched two hours or less. And girls in the study became far less active as they reached their mid-teen years.

Getting Fatter: TV and Kids

Hours of Television Watched per Day	Boys (BMI)	Girls (BMI)
<2	~21	~24
2-3	~23	~26
≥4	~25	~29

Is the Tube to Blame?

Internet

Figure 2: Case Study Screen Shot: TV and Obesity

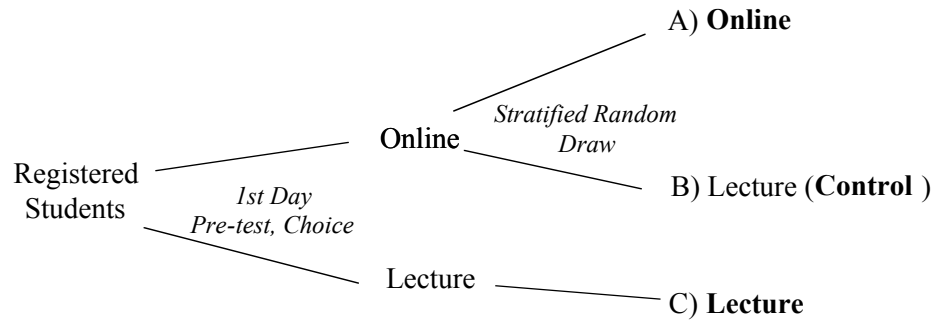


Figure 3. Semi-Randomized Design for Experiments in Year 1

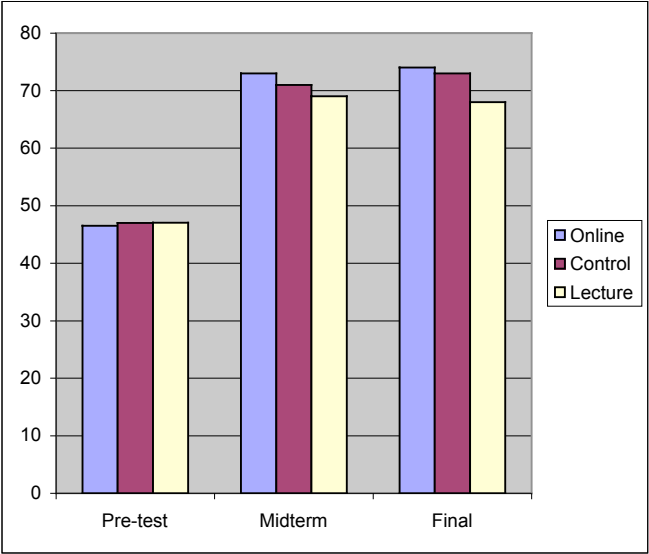


Figure 4. Winter Quarter - Year 1 (N = 180)

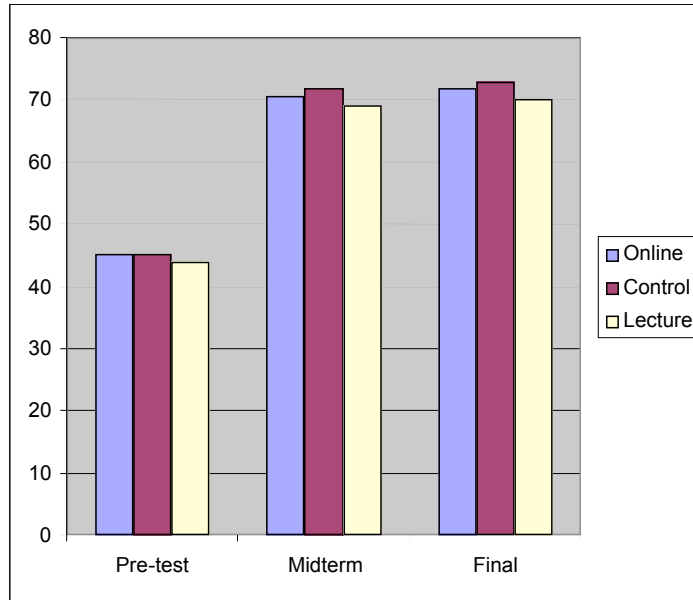


Figure 5: Spring Quarter - Year 1 (N = 130)

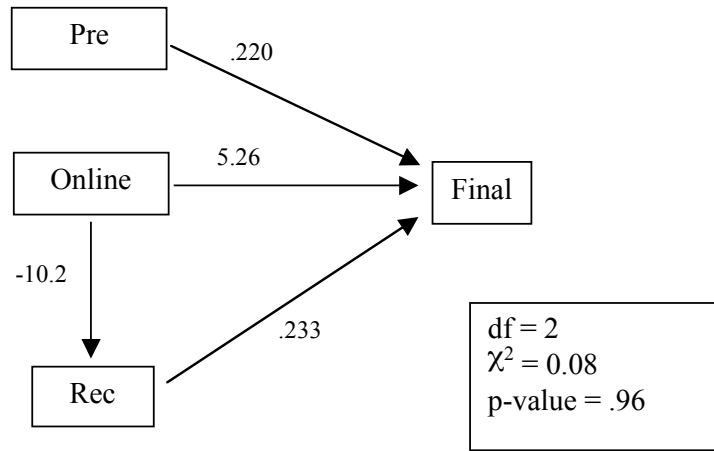


Figure 6: Two Paths from Online to Final

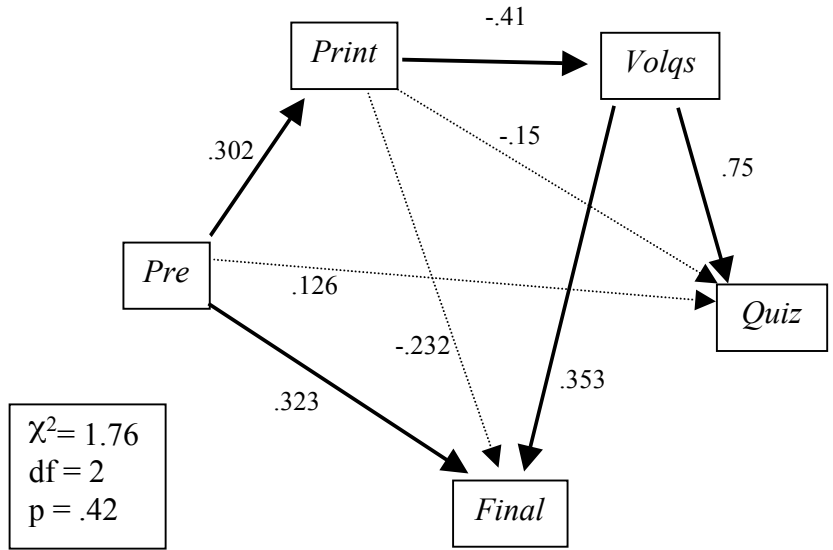


Figure 7: Best Path Model (marginally significant edges dashed)

Table 1: Year 2: Online vs. Lecture

Percentage Difference in Means: Online - Lecture				
	Pre-test	Midterm 1	Midterm 2	Final Exam
Winter (N = 157)	1.9 (p = .51)	3.6 (p = .09)	-0.35 (p = .866)	0.4 (p = .801)
Spring (N = 121)	3.3 (p = .26)	**9.8 (p < .001)	**11.2 (p < .001)	*6.08 (p = .014)

Table 2: Correlations, Means, SDs (N = 83)

	<i>Pre</i>	<i>Online</i>	<i>Rec</i>	Means	SD
<i>Pre</i>				33.83	13.15
<i>Online</i>	0.023			-	-
<i>Rec</i>	-0.004	*-0.255		78.45	19.51
<i>Final</i>	*0.287	0.182	*0.297	70.23	11.14

Table 3: Correlations, Means, SDs (N = 52)

	<i>Pre</i>	<i>Print</i>	<i>Vol</i>	<i>Quiz</i>	<i>Final</i>	Mean	SD
<i>Pre</i>	1.000					22.2/30	16.6
<i>Print</i>	0.301*	1.000				0.5	0.60
<i>VolQs</i>	-0.258	-0.421*	1.000			0.4	0.30
<i>Quiz</i>	-0.112	-0.419*	0.774*	1.000		0.5	0.20
<i>Final</i>	0.164	-0.259	0.346*	0.399*	1.000	0.75	0.10

Table 4: Predictors of Final Exam Score

Predictor	Coef	SE Coef	t	p-value
<i>Pre</i>	0.323	0.136	2.38	0.022
<i>Print</i>	-0.227	0.144	-1.57	0.122
<i>Volqs</i>	0.353	0.142	2.48	0.017

Table 5: Predictors of Quiz score

Predictor	Coef	SE Coef	t	p-value
<i>Pre</i>	0.126	0.094	1.33	0.189
<i>Print</i>	-0.148	0.100	-1.47	0.147
<i>Volqs</i>	0.750	0.099	7.48	0.000

Table 6: Alternative Models

Edge Removed from Figure 7	df	χ^2	p-value
<i>Pre</i> → <i>Quiz</i>	3	3.61	0.31
& <i>Print</i> → <i>Quiz</i>	4	5.09	0.28
& <i>Print</i> → <i>Final</i>	5	7.66	0.18