

Watch Mr. Software

Christopher A. Bohn*

Department of Systems & Software Engineering Management
Air Force Institute of Technology
Wright-Patterson AFB OH, USA
christopher.bohn@afit.edu

Abstract

We describe a distance-learning (DL) experiment to test whether interaction between the instructor and a student in the recording studio can improve the learning of remote students who do not have the opportunity to interact with the instructor during the lecture. The experiment was conducted for two software engineering continuing education courses. The in-studio students were students who had enrolled in the DL course and who also happened to work near our campus. We measured results by comparing the students' end-of-course surveys and by comparing the students' assessment scores. Our results of this initial experiment do not support the hypothesis, though we continue to investigate the question.

Keywords: Distance Learning, Continuing Education.

1 Introduction

The Air Force Institute of Technology (AFIT) conducts a series of software engineering professional continuing education courses through instructor-led distance learning (DL). The typical format is a four-week course with two hours of recorded lecture per week, supplemented with one live teleconference per week and discussion boards.

Inspired by a television marathon of “Watch Mr. Wizard” [2], we designed an experiment to determine if AFIT’s DL courses could be improved. “Watch Mr. Wizard” was a television show that taught scientific concepts its young audience through the conceit that neighborhood children visit the home of a local scientist to learn those very concepts through hands-on experiments. As the on-set child asked and answered questions, made discoveries, and voiced their insights, the chil-

dren in the television audience would also learn.

We hypothesized that including a student in the studio with the instructor when recording lectures would improve the learning of the remote students due to a combination of factors: the in-studio student’s insights would help clarify material for some students, their questions would help clarify material that hadn’t been presented clearly, and the interaction between the instructor and the in-studio student would help hold the remote students’ attention.

2 Experimental Design

Two courses were chosen for the experiment: CSE 487, Fundamentals of Object-Oriented Systems, and CSE 488, Object-Oriented System Modeling. CSE 487 provides a language-neutral introduction to object-oriented (OO) concepts: classes and objects, objects as collaborators in a decentralized solution, abstraction, encapsulation and information hiding, aggregation, inheritance and polymorphism, and OO analysis through use cases and class-responsibility-collaborator cards. CSE 488 covers the various diagrams in the Unified Modeling Language (UML) standard, how concepts elicited during analysis & design can be represented in the UML diagrams, the ways the diagrams work in concert, and how the diagrams can reveal inconsistencies in the design.

2.1 Controlled Factors

For CSE 487, the in-studio students were two recent college graduates employed at the Air Force Research Laboratory. One of the CSE 487 in-studio students continued with CSE 488; the other CSE 488 in-studio students were two practicing systems engineers.

The remote students for each class were divided roughly in half. The lectures were recorded once without in-studio students and made available to the control group, and the lectures were recorded

*The views expressed in this article are those of the author and do not reflect the official policy or position of the Air Force, Department of Defense, or the U.S. Government.

again with in-studio students and made available to the experimental group. The two groups commingled in the teleconferences and the discussion board, though the experiment was not discussed in these forums. The control group was told there was an experiment being conducted, though they were not told the exact nature of the experiment, nor that they were the control group. Obviously, it was impossible to hide the nature of the experiment from the experimental group.

We measured the results two ways. The first was the students' end-of-course surveys; the Likert-scale questions and free-answer questions were used to gauge the students' impression of the course offering. The second way we measured the results was by examining the students' homework and exam scores; their scores were used to assess the degree to which the students had learned the material.

A third batch of students later contributed to the experiment. These students were not our traditional DL students; rather, they are resident graduate students at AFIT. They had enrolled in a graduate systems engineering architecture course which pre-required the introductory software engineering course, primarily for familiarity with object-oriented concepts and with UML. The instructor agreed to let the students take an "OO Primer" as a substitute for the quarter-long software engineering course; this primer consisted of re-used lecture recordings from CSE 487 and CSE 488.

These students differed from the first two batches of students in four additional ways. First, they were not provided a copy of the textbooks for CSE 487 and CSE 488; as such, their end-of-course surveys had some questions elided. Second, they were not evaluated through homework assignments or exams; as such, they did not contribute to our second measure of the experiment. Third, they took the OO Primer during the winter break between quarters, rather than over the course of several weeks. Fourth, they had no interaction with the DL instructor.

2.2 Uncontrolled Factors

There are three factors that we did not control and that must be considered when drawing any conclusions from our experiment. First and foremost is the small sample size. Not including the in-studio students, there were nine students in CSE 487, twelve students in CSE 488, and seven students who took the OO Primer. We do not claim to have statistical significance; rather, our results must be considered to be anecdotally suggestive.

Second is that the in-studio students were given

no script; that is, they had the full freedom to offer any insight or to ask any question about something they didn't understand. We made a deliberate choice not to use "students" already familiar with the material or to provide the students with a script, so as to maximize the opportunity for them to be a "typical" student who may not understand something we thought was presented well. Further, the students were selected based on availability rather than on personality. The lack of a script, combined with the students' personalities, led to the situation that while they *could* speak as a typical student, they rarely took the opportunity to do so. Of the four in-studio students, only one was an active student; two of the students were so passive that they seldom spoke except in response to a leading question.

The third factor is that based on the time-stamps of the OO Primer surveys, those students completed both the CSE 487 and the CSE 488 surveys at the same time, rather than completing the CSE 487 survey after the CSE 487 lectures and the CSE 488 survey after the CSE 488 lectures. As such, the two surveys for the OO Primer students may not be independent measures.

3 Results

As stated previously, we measured the results two ways. We used the end-of-course surveys to gauge the students' impression of the course offering, and their homework and exam scores were used to assess the degree to which the students had learned the material.

3.1 End-of-Course Surveys

The complete end-of-course surveys are available elsewhere [1]; the Likert-scale responses are summarized in Tables 1–4.

On casual examination, our hypothesis would seem to be supported by the first batch of students (Table 1). With the exception of "How appropriate were the methods used to evaluate your contribution and learning?", the student's mean response to every question was higher for the experimental group than the control group's; similarly, for five of the nine questions, the median response was higher for the experimental group. One student's free-form answers included a comment relevant to the experiment: "I liked the concept of having students with the instructor as he gave the lecture as the interaction between the instructor and students was helpful." The only other comment that could be related to the experiment came from the control

Table 1: End-of-course survey results for CSE 487.

		Control Group					Experimental Group				
		# students	4	# students	5	# responses	3	# responses	5		
		Outstanding	Excellent	Satisfactory	Marginal	Unsatisfactory	Outstanding	Excellent	Satisfactory	Marginal	Unsatisfactory
<i>Instr.</i>	1. Instructor Effectiveness	1	1	1	0	0	2	2	1	0	0
	2. Command of Subject Matter	1	1	1	0	0	3	1	1	0	0
<i>Materials</i>	3. Course Materials Effective	1	1	1	0	0	3	1	1	0	0
	4. Materials Useful in Future	1	1	0	1	0	3	1	1	0	0
	5. Appropriate Evaluations	1	1	1	0	0	1	3	1	0	0
<i>Value</i>	6. Meaningful, Useful Info.	1	1	1	0	0	2	2	1	0	0
	7. Rate Course to Others	1	1	1	0	0	3	1	1	0	0
<i>Suprt.</i>	8. Registration	2	0	1	0	0	4	0	1	0	0
	9. Fulfilled School's Mission	1	1	1	0	0	4	0	1	0	0

Table 2: End-of-course survey results for CSE 487 portion of the OO Primer.

		Control Group					Experimental Group				
		# students	3	# students	4	# responses	2	# responses	2		
		Outstanding	Excellent	Satisfactory	Marginal	Unsatisfactory	Outstanding	Excellent	Satisfactory	Marginal	Unsatisfactory
<i>Instr.</i>	1. Instructor Effectiveness	0	2	0	0	0	0	1	1	0	0
	2. Command of Subject Matter	2	0	0	0	0	0	2	0	0	0
<i>Materials</i>	3. Course Materials Effective	0	2	0	0	0	0	2	0	0	0
	4. Materials Useful in Future	not asked									
	5. Appropriate Evaluations	not asked									
<i>Value</i>	6. Meaningful, Useful Info.	0	1	1	0	0	0	1	1	0	0
	7. Rate Course to Others	0	2	0	0	0	0	1	1	0	0
<i>Suprt.</i>	8. Registration	not asked									
	9. Fulfilled School's Mission	1	1	0	0	0	0	2	0	0	0

Table 3: End-of-course survey results for CSE 488.

		Control Group					Experimental Group				
		# students	6	# students	6	# responses	3	# responses	2		
		Outstanding	Excellent	Satisfactory	Marginal	Unsatisfactory	Outstanding	Excellent	Satisfactory	Marginal	Unsatisfactory
<i>Instr.</i>	1. Instructor Effectiveness	1	1	1	0	0	0	0	1	1	0
	2. Command of Subject Matter	1	2	0	0	0	0	1	1	0	0
<i>Materials</i>	3. Course Materials Effective	1	0	2	0	0	0	0	1	1	0
	4. Materials Useful in Future	1	1	1	0	0	0	0	1	1	0
	5. Appropriate Evaluations	1	2	0	0	0	0	0	2	0	0
<i>Value</i>	6. Meaningful, Useful Info.	1	2	0	0	0	0	1	1	0	0
	7. Rate Course to Others	1	1	1	0	0	0	0	2	0	0
<i>Suprt.</i>	8. Registration	3	0	0	0	0	0	1	1	0	0
	9. Fulfilled School's Mission	1	2	0	0	0	0	0	2	0	0

Table 4: End-of-course survey results for CSE 488 portion of the OO Primer.

		Control Group					Experimental Group				
		# students	3	# students	4	# responses	2	# responses	2		
		Outstanding	Excellent	Satisfactory	Marginal	Unsatisfactory	Outstanding	Excellent	Satisfactory	Marginal	Unsatisfactory
<i>Instr.</i>	1. Instructor Effectiveness	1	1	0	0	0	0	2	0	0	0
	2. Command of Subject Matter	1	1	0	0	0	1	0	1	0	0
<i>Materials</i>	3. Course Materials Effective	0	2	0	0	0	0	1	1	0	0
	4. Materials Useful in Future	not asked									
	5. Appropriate Evaluations	not asked									
<i>Value</i>	6. Meaningful, Useful Info.	0	2	0	0	0	0	1	1	0	0
	7. Rate Course to Others	0	2	0	0	0	0	1	1	0	0
<i>Suprt.</i>	8. Registration	not asked									
	9. Fulfilled School's Mission	0	2	0	0	0	0	2	0	0	0

group: “The material is a hard subject, had I not seen it before, I would have been lost in this class.”

The second batch of students (Table 3), though, failed to support the hypothesis. For every question, without exception, the mean response of the experimental group was lower than that of the control group; moreover for every question the experimental group’s high response was lower than that of control group, and the experimental group’s low response was lower than the control group’s for eight of the questions. Even the questions that should have been independent of the experiment (“How satisfied were you with the process you used to register for this class?” and, to a lesser degree, the “appropriate evaluations” question) showed a marked difference between the experimental group and the control group suggests the experimental group truly had a bad experience with the course. The comments from both groups were roughly the same; it wasn’t clear from the free-form answers alone that one group had a better experience than the other.

The third batch (Tables 2 & 4) also did not support the hypothesis, though not as starkly as the second batch. For both the CSE 487 and CSE 488 portions of the OO Primer, the experimental and control groups’ responses either agreed or the control group had a higher average response of a half-unit for all but one question; for “How well did the instructor demonstrate command of the subject matter?”, there was a full-unit difference between the control group’s and the experimental group’s responses. Two comments stand out as far as the experiment is concerned; one student, in the CSE 488 portion, commented, “I am not sure that the students in the lecture are very effective, no value added”. Another student from the experimental group provided a comment that may have biased his responses, “I do not like online courses”.

Treating the numbers as an aggregate (10 course-students in the control groups; 11 course-students in the experimental groups), we get Table 5. What differences there are in the high, median, and low responses indicate greater satisfaction from the control groups than the experimental groups. Among the mean ratings, the only mean response that supports the hypothesis is a $\frac{3}{125}$ -unit greater average response for the experimental group, to the question “How useful will the texts and handouts be to you in the future?” Bearing in mind how small our sample size is, this cannot be considered significant.

3.2 Student Assessment Scores

Figures 1–3 compare the CSE 487 homework, midterm exam, and final exam scores, respectively,

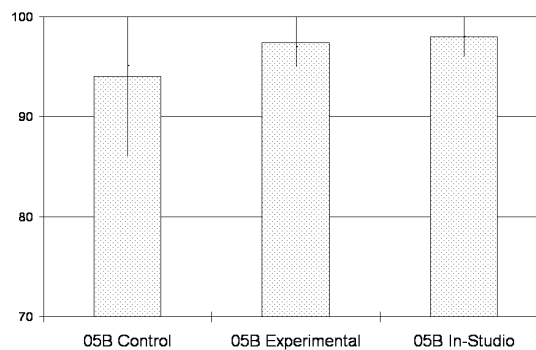


Figure 1: Homework scores for CSE 487.

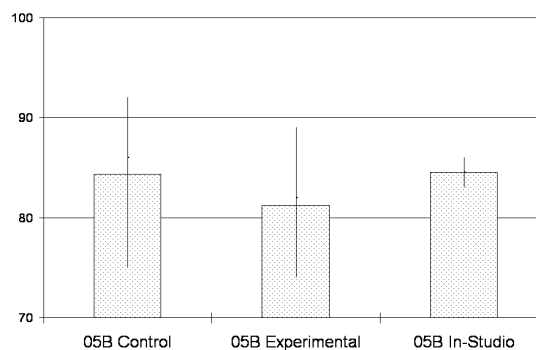


Figure 2: Midterm exam scores for CSE 487.

of the control and experimental groups, and the in-studio students. The columns indicate the mean scores; the error bars indicate the high and low scores; and the ticks on the error bars indicate the median scores. Figure 4 compares the weighted total scores for CSE 487. The scores from the homework assignments and the final exam support the hypothesis, though the scores for the midterm exam runs counter to the hypothesis. The experimental group had a higher mean, median, and low course score, but a lower high score; the differences are not statistically significant.

Similarly, figures 5–6 compare the CSE 488 homework and final exam scores, and figure 7 compares the CSE 488 weighted total scores. At first glance, it would appear the control students scored much better on their homework assignments than did the experimental students; however, this is misleading. One student in each group performed abysmally on half of their homework assignments (along the line of providing a diagram that vaguely resembled an activity diagram when a sequence diagram was specified in the problem statement). Each of these students was given the opportunity to re-accomplish the unsatisfactory assignments; only the student in the control group chose to do

Table 5: Aggregate end-of-course survey results. Likert scale: “Outstanding”=5... “Unsatisfactory”=1. Negative values for “difference” support the hypothesis.

		Control Group responses 10 for Q4,5,8 6				Experimental Group responses 11 for Q4,5,8 7				Difference			
		Mean	Median	High	Low	Mean	Median	High	Low	Mean	Median	High	Low
<i>Instr.</i>	1. Instructor Effectiveness	4.10	4	5	3	3.73	4	5	2	0.37	0	0	1
	2. Command of Subject Matter	4.40	4.5	5	3	4.09	4	5	3	0.31	0.5	0	0
<i>Materials</i>	3. Course Materials Effective	3.90	4	5	3	3.82	4	5	2	0.08	0	0	1
	4. Materials Useful in Future	3.83	4	5	2	3.86	4	5	2	-0.02	0	0	0
	5. Appropriate Evaluations	4.17	4	5	3	3.71	4	5	3	0.45	0	0	0
<i>Value</i>	6. Meaningful, Useful Info.	4.00	4	5	3	3.82	4	5	3	0.18	0	0	0
	7. Rate Course to Others	4.00	4	5	3	3.82	4	5	3	0.18	0	0	0
<i>Suprt.</i>	8. Registration	4.67	5	5	3	4.29	5	5	3	0.38	0	0	0
	9. Fulfilled School’s Mission	4.20	4	5	3	4.09	4	5	3	0.11	0	0	0

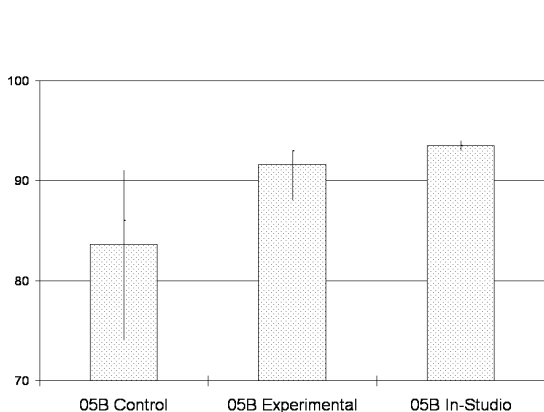


Figure 3: Final exam scores for CSE 487.

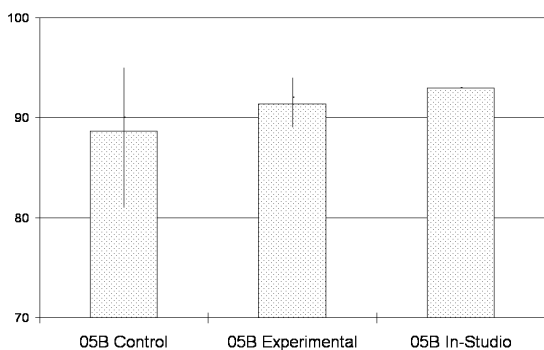


Figure 4: Weighted total scores for CSE 487.

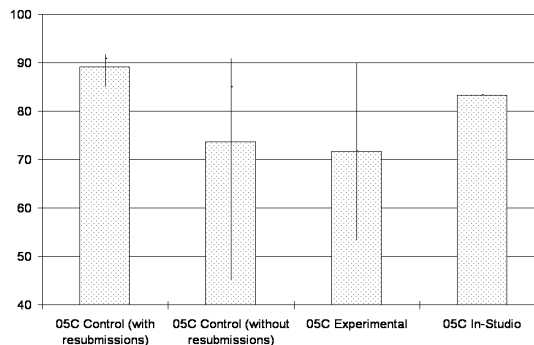


Figure 5: Homework scores for CSE 488.

so. When comparing their scores without factoring the re-accomplished assignments, the two groups each performed about the same on their homework assignments. On the final exam, while the control group did have the lowest score (the student who had trouble with his homework assignments), the remaining scores in the control group were higher than the high score from the experimental group. As we might expect from the above data, the control group fared much better than the experimental group on the weighted course score, though only slightly better after factoring-out re-accomplishment of homework assignments.

4 Conclusions and Future Work

The results of our experiment do not support our hypothesis that interaction between the instructor

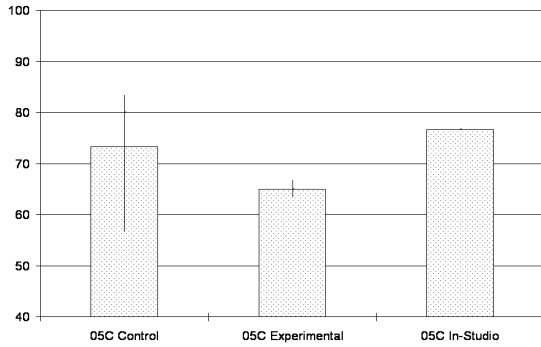


Figure 6: Final exam scores for CSE 488.

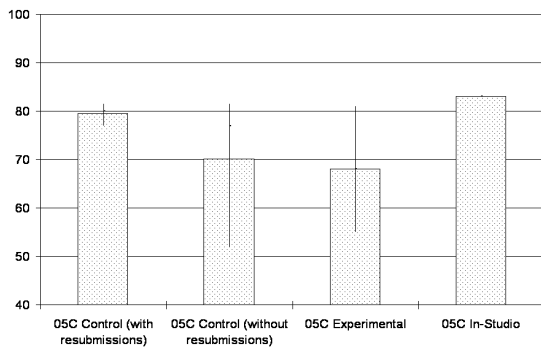


Figure 7: Weighted total scores for CSE 488.

and an in-studio student would improve remote students' learning. The students' end-of-course surveys, in aggregate, do not indicate the experimental groups had a better experience than the control groups; the groups' assessment scores were not different with any reasonable confidence. Neglecting the statistical insignificance, and ignoring experimental errors due to student aptitude and prior knowledge, we were tempted to believe that an in-studio student might be helpful for some subjects but is not helpful (and may even be harmful) for others; however (again ignoring statistical insignificance), the third batch of students' results do not support even that belief.

We have four possible explanations for this result. The first, obvious, explanation is that interaction between the instructor and in-studio students do not facilitate the remote students' learning. The other explanations are less absolute. It's possible that some students do benefit, but with the small sample size, our remote students were predominantly students who don't benefit from the in-studio student. The in-studio students' personality could also have contributed: as most of our in-studio students were "quiet" students, they may not have provided enough interaction to benefit the

remote students. Finally, it's possible that the instructor's personality was a factor: his teaching style may have been tuned to distance learning, thereby not giving the in-studio students much opportunity to offer added value.

Of course the true measure of learning isn't how the students rate the course, or even how well they performed on homework assignments and exams. More important is how well the students can apply what they learned. Toward the end of measuring how well the students can apply the CSE 487 and CSE 488 material, we have follow-on experiments in progress. For the third batch of students, those who took an OO Primer as a pre-requisite for a graduate systems engineering architecture course, we shall compare how the control and experimental groups performed in the architecture course, both against each other and against the students who had taken the standard prerequisite course. This, naturally, would have no more statistical significance than the results presented here.

Another follow-on experiment involves several dozen undergraduate students at the US Air Force Academy. These students are enrolled in a resident introductory systems engineering course and are being assigned selected lectures from the OO Primer to review as part of class preparation. As with the graduate students, we shall compare how the control and experimental groups perform in the systems engineering course, with a particular focus on the OO-related content.

References

- [1] Christopher A. Bohn. Watch Mr. Software. Technical Report AFIT/LS-TR-06-01, Air Force Institute of Technology, 2006.
- [2] The Internet Movie Database. "Watch Mr. Wizard" (1951). <http://www.imdb.com/title/tt0043244/>. Viewed 5 March 2006.