

Examining the Effectiveness of
Blended Instruction on Teaching
Introductory Statistics

Mahtash Esfandiari

Chris Barr

Adam Sugano

UCLA Department of Statistics May
2006

I Background of the study:

The old four-unit Statistics 10 was restructured such that it is now a five-unit GE approved course in the Foundations of Scientific Inquiry area, both in the Life Sciences and Physical Sciences subgroup.

I.1 The major goals of the restructured course are to:

- Maximize the role of students as active learners,
- Introduce statistics as a “science of data”,
- Create an interactive environment, and
- Place the major emphasis on upper level thinking and problem solving.

I.2 Blended instruction, the use of technology and regular methods of teaching, was implemented to reach the above goals:

- A software called Moodle is adapted that makes it possible for the students to take on-line quizzes and provide the students and the instructors with information regarding correct answers and summary statistics.
- Students take weekly on line quizzes prior to attending section.
- In section, they are placed in groups (based on quartiles) to discuss the problems that the majority got wrong. The groups stay intact for the whole quarter.
- They take the quiz for the second time after group discussion,
- They do a weekly lab.

I.2 Blended instruction, the use of technology and regular methods of teaching, was implemented to reach the above goals:

- They have two hours of lecture with the instructor (instead of three) and two hours with the teaching assistant (instead of one).

II Piloting of the restructured Statistic 10

The restructured Statistics 10 was piloted in Winter 2005. Based on this exploratory case study, we succeeded in reaching the overall goals of the study.

Starting Fall 2005, all of the sections of Statistics 10 are taught through “blended instruction”. About 2000 students take this course per academic year.

III Description of the study:

III.1 Description of the design

An experimental study was designed to investigate the effectiveness of “blended instruction” on teaching introductory statistics to large groups of students (100 or more). The control group consisted of a large group of students that were taught statistics 10 by the old method in Spring 2005. The experimental group consisted of a similar group of students that were taught the restructured statistics 10 by the same instructor in the Fall of 2005.

III.2 The similarities between the experimental and the control group

- The time of the day the class met,
- The instructor,
- The book and lecture notes,
- Method of teaching implemented by the instructor (generative method),
- The type of thinking the students were expected to engage in,
- Sample quizzes and review midterm,
- Examples using real data sets and computer printouts,
- Weekly homework, and
- Similar midterm and the final examination. (Students do not get the final back and thus it was fine to use the same exam for both groups.)

III.3 The differences between the experimental and the control group

The experimental group:

- Took weekly on-line quizzes.
- Met in groups of four to discuss the weekly on-line quizzes.
- Did a weekly lab.
- Had two hours of lecture with the instructor.
- Had two hours with the TA.

The control group:

- Did not take on-line quizzes.
- Did not do a lab.
- Had three hours with the instructor.
- Had one hour with the TA.

III.3 The differences between the experimental and the control group continued:

- The control group received four mock quizzes, The objective was to give the students a chance to independently think about the concepts taught in the course. The students were expected to look through the mock quizzes prior to lecture. Whenever possible, the mock quizzes were discussed during lecture and the students were asked to discuss the questions with their neighbor. However, they were not graded and so we could not check whether they did the mock quizzes on their own or not.
- For the experimental group, the mock quizzes were turned into weekly on-line quizzes. As elaborated before the weekly on-line quizzes were quite extensive and in-depth. They required a lot of individual and group interaction on the part of the students and the mock quizzes did not.

III.4 Demography of the participants

Table 1: The percentage of male and female by group

	% Male	% Female
Control	36.6	63.4
Experimental	44.3	55.7

Table 2. The percentage of freshman, sophomore, junior, and senior by group

	% Freshman	% Sophomore	% Junior	% Senior
Control	50	29.7	12.3	8
Experimental	69.7	12	10.3	8

III.4 Demography of the Participants continued

Table 3. Percentage of whether statistics 10 was a graduation requirement by group

	Graduation Requirement	Not a Graduation Requirement
Control	76.3	23.7
Experimental	65.9	34.1

Table 4: Students' GPA by group

	GPA*
Control	3.27
Experimental	3.21

The data was collected on the day of the final examination. The freshmen were asked to estimate their GPA based on the midterm grades.

IV Objectives of the study:

Examining the impact of blended instruction on

- 1) Attendance and homework:
 - The number of hours the student attended lecture.
 - The number of hours the students attended section.
 - The number of hours the students spent doing homework

- 2) The acquisition of statistical knowledge as measured by multiple choice tests and open-ended questions.

IV Objectives of the study continued:

- 3)** Upper level thinking including application, analysis, synthesis, and evaluation.
- 4)** Introducing statistics as a “science of data”,
- 5)** Enhancement of generative teaching,
- 6)** Helping students to be active learners and generate their own knowledge.

V Operational definition of major variables of the study

V.1 Generative teaching is defined as creating an environment to:

- Motivates the students,
- Gets their attention,
- Minimizes their role as passive recipients of information,
- Maximizes their roles as active learners,
- Help them create a tie between the old and the new information, and
- Help them generate their own knowledge.

V.2 Upper level thinking is defined as conducting teaching and assessment in such a way to:

- Minimize memorization, and
- Maximize Problem solving and upper level thinking including application, analysis, synthesis, and evaluation.

V.3 Introducing statistics as a science of data is defined as:

Introducing statistic not as a branch of mathematics and stepwise calculation without any context but as a mean of,

- understanding and interpreting data,
- answering real world questions with a real context, and
- understanding the data and findings reported in the mass media.

V.4 An interactive environment is defined as a setting that:

- Allows the students to interact with each other and play an active role in clarification and understanding the concepts and methods that the majority of students found to be vague based on formative evaluation (i.e. weekly quizzes).
- Allows the instructor and the teaching assistant to play the role of a facilitator and guide the students to find the answers to the problems rather than providing them with the answers.

VI Conclusions:

VI.1 Effect of blended instruction on the percentage of time students attended lecture

Blended instruction did not have any effect on the percentage of time that students attended lecture ($t = 0.686$, $P = 0.494$)

Table 5: Mean, and standard deviation of the average percentage of time that the experimental and the control students attended lecture

Group	N	M(SD)	t	P value
Control	82	80.25(24.63)	0.686	0.494
Experimental	87	77.45(28.00)		

VI.2 Effect of blended instruction on the percentage of time students attended section

Blended instruction had a positive and statistically significant effect on the average percentage of time that the students attended section ($t = 8.263$, $P = 0.000$). The average percentage of time that the students attended section was 46.48% higher for the experimental group (92.54% Vs. 50.06%).

Table 6: Mean, and standard deviation of the average of time that the experimental and the control students attended section

Group	N	M(SD)	t	P value
Control	81	56.04(36.73)	8.47	0.000
Experimental	87	92.53(15.73)		

VI.3 Effect of blended instruction on the time students spent on homework

Blended instruction had no effect on the number of hours that the students spent on homework per week ($t = 0.043$, $P = 0.966$).

Table 7: Mean, and standard deviation of the average time that the experimental and the control students spent on homework

Group	N	M(SD)	t	P value
Control	78	2.32(1.18)	0.043	0.966
Experimental	82	2.33(1.36)		

VI.4 Comparison of the performance of the control and experimental students on multiple-choice questions

The same final examination was administered to both the control and the experimental group.

- **The experimental students performed better on complex questions that involved analysis and synthesis of information. (See Figure 1)**
- **The control students performed slightly better on questions that involved definitions or recall of information. (See Figure 2)**

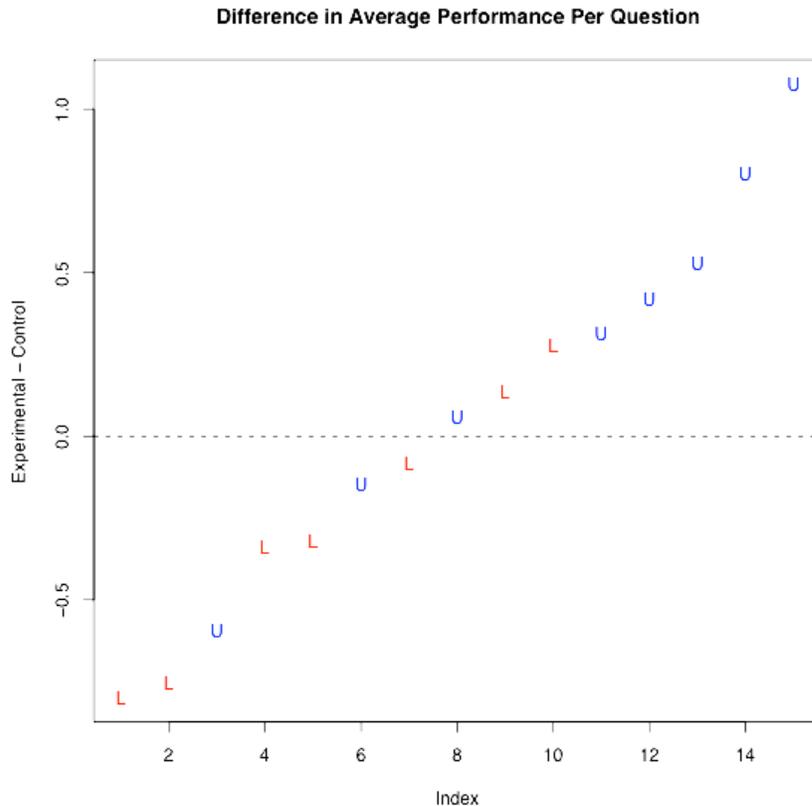


Figure 1: Relative performance of the control and experimental students on multiple-choice questions that involved upper level (U) and lower level thinking (L)

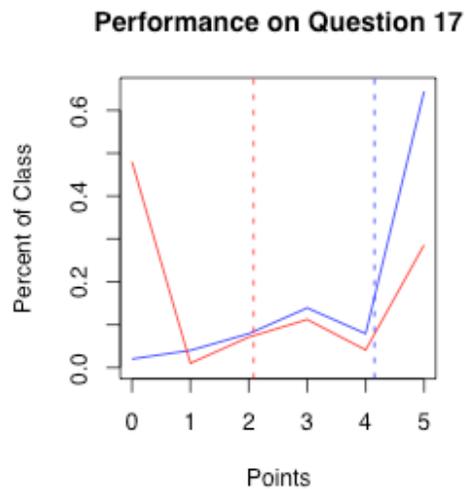
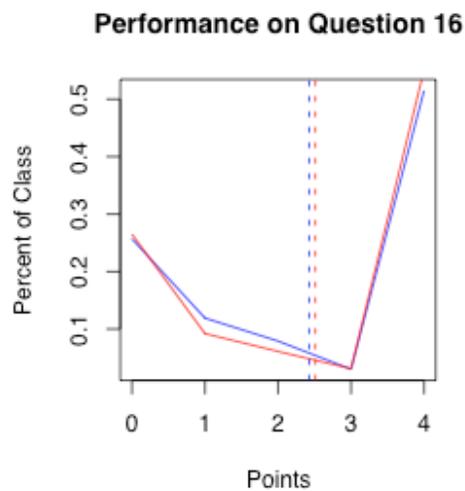
VI.5 Comparison of the performance of experimental and control students on open-ended questions

- The students in the experimental and control group did equally well on the free response question that related to hypothesis testing and involved calculation and interpretation of the findings. (see Figure 2)**
- The students in the experimental group performed much better on the open-ended question that related to Central Limit Theorem and involved analysis and evaluation. The students in the control group just used the right statistical terminology. The students in the experimental group had the right idea and could elaborate what the statistical terminology meant. (see Figure 3)**

Figures 2 and 3: Relative performance of the control and experimental groups on open-ended questions

Question 16 involved calculation

Question 17 involved analysis and evaluation



VI.6 Effect of blended instruction on the students' perceptions regarding the extent to which they engaged in upper level thinking

No statistically significant difference was found in the perceptions of the experimental and the control students with respect to their involvement in upper level thinking ($t = 1.67, P = 0.11$).

Table 8: Mean, and standard deviation of the students' perceptions regarding their involvement in upper level thinking

Group	N	M(SD)	t	P value
Control	70	57.00(18.07)	1.61	0.110
Experimental	76	52.13(18.53)		

VI.7 The effect of “blended instruction” on students’ perceptions regarding the extent to which statistics was introduced as a “science of data”

No statistically significant difference was found between the perceptions of the experimental and the control students with respect to how well we had succeeded in introducing statistics as a science of data ($t = 0.371$, $P = 0.711$). This result is expected; because although the control group students did not do a lab, through the weekly handouts they were exposed to the application of computers in statistical analysis and interpretation of the resulting outputs.

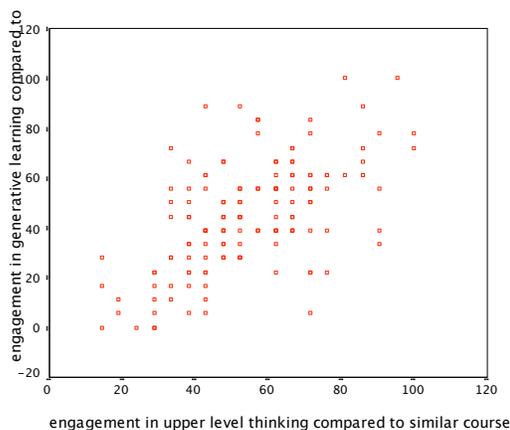
Table 9: Mean, and standard deviation of the students’ perceptions regarding the extent to which statistics was introduced as a science of data

Group	N	M(SD)	t	P value
Control	73	61.55(20.86)	0.371	0.711
Experimental	83	60.24(23.01)		

VI.8 The relationship between students' perceptions of engaging in upper level thinking and generative learning

A positive and significant correlation was found between students' perceptions of their engagement in upper level thinking and generative learning ($r = 0.583$, $P = 0.000$). This finding was true for both groups { r (control) = 0.599, $P = 0.000$, r (experimental) = 0.543, $P = .000$ }. This finding indicates that the majority of the students who were above average on upper level thinking were also above average on generative learning.

Figure 4: Scatter plot of the students' perceptions of their involvement in upper level thinking and generative learning



VI.9 The effect of “blended instruction” on students’ perceptions of the generative nature of lecture and the generative nature of homework

Multivariate analysis of variance (MANOVA) was conducted to compare the experimental and the control students’ perceptions of the generative nature of the lecture and the homework. Results indicated that there was a statistically significant difference between the perceptions of the experimental and the control group regarding the generative nature of the lecture ($F = 21.589, P = 0.000$) but not on homework ($F = 0.048, P = 0.854$).

Univariate tests indicated that the control group perceived the lecture to be more generative in nature than the experimental group ($F = 39.410, P = 0.000$). This result is not surprising because in the case of the control group the generation of knowledge happened mainly in lecture which consisted of one more hour of weekly contact with the instructor as well as discussion of mock quizzes with the instructor and with the

student next to them.

VI.10 The effect of “blended instruction” on students’ perceptions of the generative nature of lecture and the generative nature of homework continued

Univariate tests indicated that there was no statistically significant difference between the perceptions of the experimental and the control group regarding the generative nature of the homework ($F = 0.047$, $P = 0.876$). This result is not surprising, because, both groups received the same homework, they spent the same number of hours on weekly homeworks (2.37 for both groups), and the percentage of points assigned to homework was the same for both groups.

Table 10. Mean, standard deviation, univariate F and P values related to the students' perceptions of the generative nature of lecture and homework by group

	Lecture	Homework
	M(SD)	M(SD)
Control	55.19(18.43)	51.77(19.00)
Experimental	35.49(19.35)	52.43(17.71)
Univariate F and P value	39.41(0.000)	0.047(0.830)

VI.11 Comparison of the experimental students' perceptions regarding the generative nature of the different sources of instruction

Repeated measures ANOVA indicated that the experimental students' perceptions of the generative nature of lecture, lab, homework, quiz, and group discussion of the quiz differed significantly ($F_{4, 61} = 26.156, P = 0.000$).

Results indicated that the factors that enhanced generation of knowledge ranked as follows:

- On-line quiz (rank 1)
- Discussion of on-line quiz (rank 2)
- Homework (3)
- Lab (4)
- Lecture (5)

Figure 5: Plot of the experimental students' perceptions regarding the generative nature of the different sources of instruction

1 = Lecture

2 – Homework

3 =Lab

4 – On-line quiz

5 = Discussion of the quiz

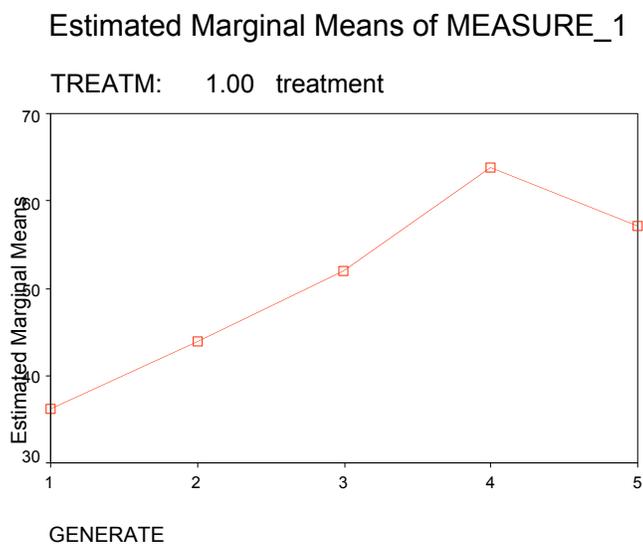


Table 11: Mean and standard deviation of the experimental students' perceptions of the generative nature of the different sources of instruction

	Mean	SD
Lecture	36.12	19.56
Lab	43.93	22.67
Homework	51.94	19.14
On-line quiz	63.81	19.32
Discussion of on-line quiz	57.17	18.97

VI.12 Conclusions regarding the students' perception of interactive nature of "blended instruction"

- **33.3%** of the experimental students indicated that **weekly meetings of the instructor** with the small groups helped to create a closer teacher-student contact to a large or a very large extent.
- **65.1%** of the experimental students indicated that the **group discussion of the on-line quiz** helped to create a closer student-student contact to a large or a very large extent.
- **64.4%** of the experimental students indicated that the **group discussion of the on-line quiz** helped to create a closer student-TA contact to a large or a very large extent.

VI.13 Comparison of the experimental students' perceptions regarding the level of interaction with other students, TA, and the instructor

Repeated measures ANOVA indicated that the experimental students' perceptions of the level of interaction that they had with other students, the TA, and the instructor differed significantly ($F_{2, 82} = 15.224, P = 0.000$).

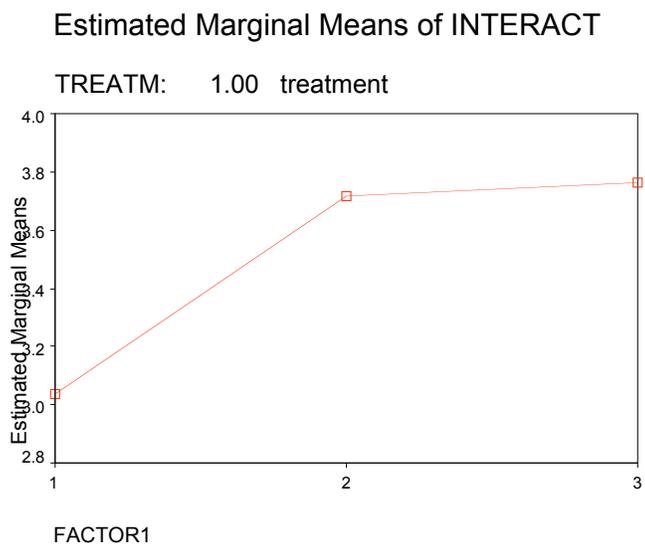
Results indicated that the level of interaction between the students themselves and the student and the TA was similar and higher than the interaction between the students and the instructor.

Figure 6: The plot of the mean for the experimental students' perception of the level of interaction between

1 = Student instructor

2 = Student-student

3 = Student - TA



VI.14 Qualitative Evaluation

Both the control and the experimental students were asked to respond to the following questions:

Question 1: What was the best feature of this course?

Question 2: How can things be improved?

The students' comments were read and classified into four categories. The results are presented in Tables 12 and 13

Table 12: Summary of students' comments regarding the best features of the old (control group) and restructured (experimental group) Statistics 10

units = %	lecture	Discussion	None	Other
Control	53	1	30	16
Experimental	12	54	21	13

Table 13: Summary of students' comments regarding what needs to be improved in the old (control group) and restructured (experimental group) Statistics 10

units = %	lecture	Discussion	None	Other
Control	20	20	37	23
Experimental	38	12	27	23

VI.15 Qualitative Evaluation Continued

As indicated by tables 12 and 13, the qualitative data are aligned with the quantitative findings. The group that was taught by “blended instruction” found section to be the best feature and the most generative dimension of the course. It is during section that they discussed the on-line quiz and the lab and interacted with each other and the TA. Whereas, the control group found the lecture to be the best feature of the course because most of the interaction and generation of knowledge happened during lecture and not discussion. During section all that the control group did was to have the TA help them with solving homework problems.

VI.15 Qualitative Evaluation Continued

The qualitative and quantitative findings reported above are aligned with classroom observations that were made of the teaching assistants that taught the old and the restructured Statistics 10. Based on these observations:

- The attendance has almost doubled. For the sections that meet at 10 and later, there is full attendance. For sections that meet at 8 and 9 AM, there is 60% to 70% attendance.
- There is a lot more interaction between the students.
- There is a lot more interaction between the students and the TA.
- The students have a more in-depth understanding of the material.
- The section is not limited to the students copying the answers to the homework problems from the board.
- The students are more engaged in generation of their own knowledge.
- The TA seems more involved and engaged.

VII Conclusions

- Upper level thinking and generation of knowledge has been enhanced.
- Student - Student and Student - TA interaction has increased. Increased interaction has been shown to have long term benefits on intellectual, personal, and professional life of individuals.
- Students feel more comfortable asking questions.
- Attendance in section has improved (almost doubled).
- TA has a better feel of student progress, more responsibility, and more opportunity to improve teaching skills.
- Discussion of on-line quizzes has provided the TAs with the opportunity to engage students in discussion.
- There is less time to cover the same amount of material.
- More thought about the software and labs.
- Work towards creating a comprehensive syllabus that can be used by any instructor.
- Use different quizzes on the two attempts
- Offer more Moodle training for the instructors and TAs.
- Edit and classify the questions in the test bank.

VIII **Recommendations**

- Revised syllabus that can be taught under the constraints of the new Statistics 10. (Done)
- Select one software for all instructors to use - Fathom is considered the best option. (In progress)
- Develop a lab manual to complement the new software. (In progress)
- Develop a buddy (peer mentoring) system for experienced TA s to support new Teaching Assistants in learning how to teach in this new format
- Have instructors and teaching assistants explain to students the motivation behind generative learning and why teaching assistants cannot provide answers directly. (in progress)
- Improve the test bank, making questions easier to navigate.
- Find a software that helps to grade open-ended questions. We plan to field test a software called CPR.

IX Acknowledgments

- College of Letters and Sciences
- Office of Instructional Development
- Participants of Focus Group
- Jan Deleeuw, the Chair, and Faculty of the Statistics Department
- Teaching Assistants at the Statistics Department
- Jose Garcia, Computing Resource Manager at the Statistics Department

X Appendix

TA FOCUS GROUP – APRIL 24TH 2006

OBJECTIVE: The objective of this focus group was to seek the input of the TAs regarding the restructured stat 10. A total of 11 TAs some of whom had taught both versions of Stat 10 and some of whom had only taught the revised version of Stat 10 participated.

I Summary of the TA comments

I.1 Comments regarding the on-line quizzes

- I like the quizzes. They help the students stay connected. They make the attendance increase.
- At the beginning I thought I had to give them the answers. Then, I observed Adam and decided that it was best to have them come up with the answers.
- Definitely a lot more interaction because of labs and quizzes. It makes it easier for me to interact with the students.
- There is more show of students.
- The students feel like they are teaching themselves (they do not necessarily see this as something positive).
- The blended instruction is a much better alternative. There is higher attendance. The lab and the quizzes give us something to talk about. Putting people in groups works very well. When they meet in their groups, I do not give them the answer. If you hold your ground, they listen. I sometimes ask one group if the answer of the other group is correct.
- Struggle is essential and important for them to make their own knowledge.
- It is not a good idea to use the same quiz for both attempts. They all try to get the right answers to the quiz.
- A lot of them know what the answer is, but, they do not know why.

I.2 Comments on group discussion of the quiz among students

- Group discussions are very helpful.
- It is helpful for me to see what the different groups are saying

II Comments on labs

- Labs are rushed. They do not think about concepts.
- You can take one lab and make it into three or four labs.
- Students do not like the lab and they complain that they do not have enough time to do the lab and the quizzes.
- The major objective of the labs should not be how to use the software, it should be how to interpret the printouts.
- This quarter we are not making any labs due till week three and this way they learn the software first.
- They like Fathom. It is more visual than STATA and they like that.
- Students like the fact that they can download R and work on it from home.
- Fathom is better. They make a lot of typos with STATA.
- Fathom is great for linking of plots.
- Labs should be very simple. Labs should not be taught technically.
- The idea behind labs is not to teach programming.

III General comments about improvement

- The one hour of lost lecture makes a lot of difference. Sometimes I feel like a mini-professor .The TA should spend time on Tuesday and Thursday trying to make sure that the students know the topic.
- There is not enough time to teach the material, especially when you have a lab too.
- I do not recommend having the section at 8 AM. Later in the day it helps.
- We cannot make up the lost hour in two hours. Time is really crunched.
- Professors' approaches are different. We need to be coherent. Professor need some kind of Moodle training.
- Better training with Moodle for the students. Tell the students to use the forum. You can reply to them through Moodle.
- You do loose something by meeting with the professor twice a week. Find those TAS who like teaching and assign them to Stat 10.

IV My observation of the restructured Stat 10 TAs

My observations are totally consistent with TA comments:

With the restructured stat 10, there is

- a) More interaction among students
- b) More interaction among TA and students
- c) The students are up to date with the relevant topics
- d) The quiz and the lab give the TAs and the students something to talk about
- e) Students seem more motivated

Based on my observations of the restructures Stat 10, I can place the TAs in three categories:

- a) Some TAs solve all the quiz problems on the board and give the students the answers.
- b) Some TAs place the students in groups and let them discuss the questions. But, they also provide them with the answers.
- c) Some TAs place the students in groups and let them decide what the right answer its. They guide them to find the right answer. But, do not give it to them.

V Suggestions:

Given that our teaching load and the number of students who are going to enroll in Stat 10 will be increasing, it does not seem practical (from an administrative point of view) to raise the number of hours of instruction from two to three. Along the same line, it seems that the original model where the instructor was expected to spend the third hour paying weekly visits to the sections does not look to be practical either. Given these facts, and the fact that the teaching assistants feel overwhelmed with having to provide help with homework, quiz review, the lab, and making sure that the students have understood the prerequisite concepts and methods, the following suggestions are made:

- Given that the objective of Statistics 10 is to introduce the students to statistics as a “science of data”, help them interpret the statistical findings reported in media, and see how statistics is used in solving real world problems, we will need to emphasize only on those concepts and methods that help us reach this objective. Thus, it is suggested to have the instructors who teach STAT 10 agree on the content that can be realistically taught within the given time frame and will help us reach the above objectives. Develop a syllabus that is acceptable to the instructors who teach the course.
- Choose a software that is easy to use and needs the least amount of programming. Based on the input of the teaching assistants, Fathom seems to be a good candidate.
- Have the teaching assistants who have taught the restructured Stat 10 agree on a number of “simple and basic labs” that the students can do with least amount of guidance and on their own. Develop a lab manual that the future teaching assistants can use.
- Develop a buddy (peer mentoring) system in which the teaching assistants who have taught Stat 10 and found it enjoyable and rewarding play the role of a coach to the new teaching assistants and tell them the “dos and don’ts”.

V Suggestions continued:

- Given that the major goal of the redesigned course is to have the students play an active role in their own learning, have the teaching assistants explain to the students why they cannot provide them with the answers to the quiz problem and have them hold their ground. Otherwise, there is no difference between copying the answers to homework problem and quiz problems. The TAs need to have this approach modeled for them.