METHODS OF LEARNING IN STATISTICAL EDUCATION: A RANDOMIZED TRIAL OF PUBLIC HEALTH GRADUATE STUDENTS

FELICITY BOYD ENDERS, PHD
Mayo Clinic, Division of Biostatistics
enders.felicity@mayo.edu

MARIE DIENER-WEST, PHD
Johns Hopkins University, Department of Biostatistics
mdiener@jhsph.edu

ABSTRACT

A randomized trial of 265 consenting students was conducted within an introductory biostatistics course: 69 received eight small group cooperative learning sessions; 97 accessed internet learning sessions; 96 received no intervention. Effect on examination score (95% CI) was assessed by intent-to-treat analysis and by incorporating reported participation. No difference was found by intent-to-treat analysis. After incorporating reported participation, adjusted average improvement was 1.7 points (-1.8, 5.2) per cooperative session and 2.1 points (-1.4, 5.5) per internet session after one examination. After four examinations, adjusted average improvement for four study sessions was 5.3 points (0.4, 10.3) per examination for cooperative learning and 8.1 points (3.0, 13.2) for internet learning. Consistent participation in active learning may improve understanding beyond the traditional classroom.

Keywords: *Statistics education research; Active learning; Cooperative learning; Internet learning; Randomized trial*

1. INTRODUCTION

The discipline of statistics provides critical quantitative tools for public health researchers and practitioners. Students pursuing graduate degrees in public health must become familiar with key concepts in statistical reasoning and knowledge of the appropriate use and interpretation of classical biostatistical methods such as estimation, hypothesis testing, and multivariable analysis. In particular, the widespread availability and accessibility of statistical computing has increased the potential for public health professionals to confront statistical analyses in published reports, perform their own data analyses, or collaborate with research teams.

Because of their quantitative nature, courses covering statistical concepts and methods may be challenging for students from other fields of study. A variety of reasons have been proposed to explain why students might have difficulty in developing introductory statistical skills and competencies. Such students frequently harbor long-held anxiety regarding mathematical courses and traditional didactic teaching methods may not allow them to sufficiently overcome such fears (Bradstreet, 1996). In addition to these barriers, students are often enrolled in multiple courses or concurrently employed, leading to a stressful background environment (Simpson, 1995). Finally, courses in introductory statistics draw such a variety of students from diverse backgrounds and with different prior knowledge and innate skills that it can be exceedingly challenging for instructors to simultaneously tailor didactic course material to meet all student needs (Simpson, 1995).

Recent advances in educational psychology and computer technology suggest possible ways to improve students' conceptual understanding of key statistical concepts. New instructional methods may enhance statistical education and students' learning of statistical concepts. One way to tailor statistical education is to include active learning methodology. "Active learning" refers to engaging a student in an activity, as contrasted with a lecture format or textbook which solely provides the student with information. A review of the literature in statistical education reveals that students may learn more readily when material is presented through student interaction or activities, as compared to the traditional passive lecturing style (Bradstreet, 1996; Garfield, 1995a; Garfield, 1995b; Lovett & Greenhouse, 2000; Moore, 1997). Ideally, this direct interplay forces students to overturn misconceptions, fears, or learning difficulties that hamper their ability to develop correct statistical intuition (Garfield, 1995a; Garfield, 1995b; Lovett & Greenhouse, 2000). Including such methodologies in the learning process might help improve students' understanding of statistical concepts. By establishing a "hands-on" environment, active learning may help alleviate difficulties heightened by anxiety related to mathematical concepts.

Active learning can be facilitated in a number of ways. "Cooperative learning" is accomplished when students work together on a structured activity in small groups to gain conceptual understanding (Garfield, 1993). This can be accomplished during, after, or instead of a traditional lecture. One method is to reinforce concepts and techniques introduced in a didactic lecture by subsequent small group activities facilitated by a teaching assistant. By working together, students not only engage in active learning, but derive benefits from their combined knowledge base.

Although the majority of previous attempts to implement active learning within statistical classrooms have used a cooperative learning approach (Gnanadesikan, Scheaffer, Watkins, & Witmer, 1997; Kvam, 2000; Magel, 1998), this might be difficult to accomplish with a large class size. Magel (1998) used cooperative learning in a class of 195 students and found it required significant advance preparation to break students into the small groups required and still have a single instructor serve as a facilitator for all the groups. Creating an interface with active learning using currently available internet technology provides an alternative approach for improving student understanding in large classes with a didactic course format. JAVA applets (mini-applications) provide a venue for students to independently examine statistical phenomena within a controlled internet-based environment. The interactive nature of the applets allows active learning to take place on the computer, i.e., "internet learning." Internet learning is distinct from "hybrid learning" (Utts, Sommer, Acredolo, Maher, & Matthews, 2003; Ward, 2004). In a hybrid course, the bulk of the course is online, and in person contact with students is limited, often to approximately an hour per week. By contrast, internet learning acts as an online component added to a traditional didactic course.

Previous studies have described the use of cooperative learning (Gnanadesikan et al., 1997; Kvam, 2000; Magel, 1998; Shaughnessy, 1977), but very few studies have compared cooperative learning or technologically-enhanced learning with the more traditional didactic or lecture-based style. This research study focuses on the implementation and evaluation of the addition of innovative instructional methods to an existing didactic course sequence in introductory biostatistics for non-statisticians. The present study was designed to evaluate cooperative learning and internet learning within a randomized setting, and to compare the relative merits of cooperative and internet learning to each other and to a control group.

2. METHODS

2.1. STUDY DESIGN

This study was conducted from September through December 2001 (16 weeks) in the context of an introductory biostatistics course that was a requirement for students in most Masters and Doctoral degree programs at a school of public health. Standard course instruction included 3 hours of lecture and one 2-hour laboratory each week. The laboratory consisted of a structured review of examples pertaining to lecture material but in a smaller group setting that permitted more discussion. The first half of the course reviewed introductory concepts such as graphing, summary statistics, exploratory data analysis, probability concepts and distributions, and estimation and hypothesis testing. The second half of the course covered inference for one or two groups, analysis of variance, and simple linear regression. Learning materials consisted of lectures, accompanying lecture notes, laboratory exercises, problem sets, online self-evaluation problems, Stata™ (The Stata Corporation, 2001) computing notes, quizzes, and examinations.

The study design was a randomization among consenting students to one of three groups: cooperative learning (in person), internet-based learning (online), and control (see Figure 1 for a schema of the study design and participation). During the first week of classes, students were offered the opportunity to participate in the study and asked to complete an online pre-study survey of their mathematical and statistical skill and aptitude as well as demographic characteristics. All students were eligible, but were enrolled in the study only after providing written informed consent. In order to ensure representation of all degree programs, the randomization was stratified by degree program (Doctoral, MPH, other Masters degree, or Other). After randomization, the intervention phase was initiated. Each of the intervention sessions began after the introduction of the relevant concepts in lecture, and followed the same basic framework. Students in the cooperative learning group attended a one hour bi-monthly small group active learning session facilitated by a single experienced teaching assistant who did not participate in any other course-related activities. Many of the active learning sessions were motivated by projects described in *Activity-Based Statistics* by Scheaffer, Gnanadesikan, Watkins, and Witmer (1996).

At the same time, students in the internet learning group individually accessed a specially designed internet learning website and completed an internet-based activity typically focused on statistical concepts illustrated by interactive JAVA applets. The website was comprised of applets publicly available on the internet that were designed to help students learn particular statistical concepts. For each session, links to these applets were embedded in a single computer screen providing short instructions and questions for the students. The applets and their instructions remained available to students throughout the study.

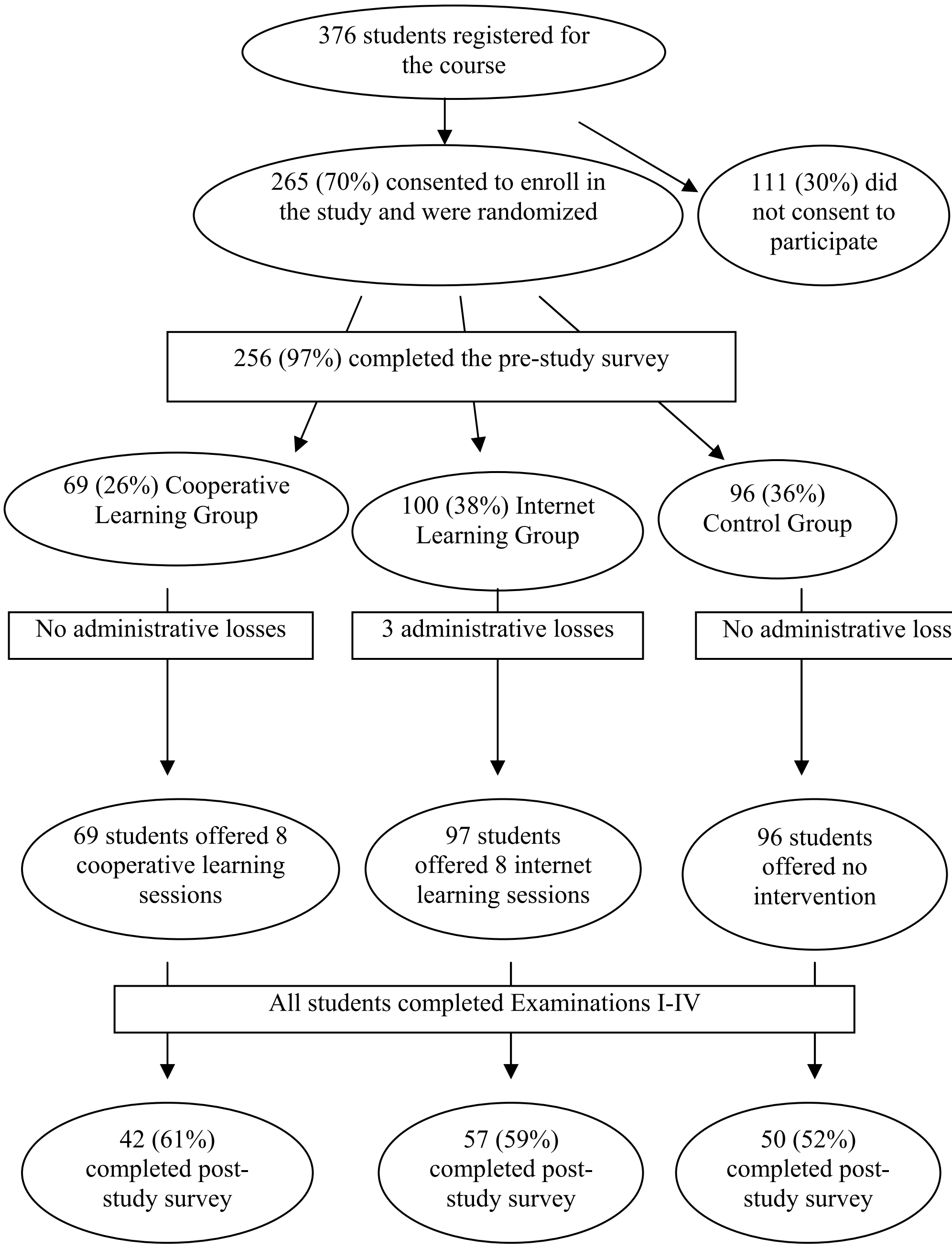

Figure 1. Study design and participation

The intervention sessions covered eight topics deemed integral to the understanding of course material: 1) conditional probability in a 2×2 table; 2) the Binomial and Poisson distributions; 3) the sampling distribution of the sample mean; 4) hypothesis testing; 5) confidence intervals; 6) the X^2 distribution; 7) analysis of variance; and 8) simple linear regression. Assessments were based on student performance as measured by four course examination scores. The first course examination was administered after the second study session; the second course examination was administered after the fourth study session; the third course examination was administered after the sixth study session; and the fourth and final course examination was administered after the eighth study session. Each examination focused on material since the prior examination and included 20 five point questions so that possible scores ranged from 0 to 100 points.

2.2. STATISTICAL ANALYSIS

The primary goal of the analysis was to investigate possible associations between intervention and student performance in the course as measured by course examination scores. Three separate linear modeling approaches were used to compare student performance by study group (McCullagh & Nelder, 1989). In the first two approaches, the four examination scores for each student (0 to 100 points) were used as a set of four longitudinal outcomes with an exchangeable covariance structure; in the third approach, the outcome was the cumulative examination score (0 to 400 points). See Appendix A for equations used in each of the three approaches.

Intent-to-treat models: The first approach utilized the intent-to-treat principle; in Model 1, examination score was regressed on assigned study group. A random effect at the student level was employed for the repeated measures structure resulting from the use of the four examination scores for each student. Three indicator variables were included to adjust for variability in scores across the four course examinations (the fourth examination served as the reference).

Individual reported participation models: The second approach incorporated students' reported participation in the sessions. In Model 2a, participation in the most recent study session in either the cooperative learning group or the internet learning group or participation in neither session was used to predict the subsequent examination score. Model 2b used participation in both of the two most recent sessions. Similarly, Models 2c and 2d incorporated participation in the three most recent sessions (if available), or the four most recent sessions (if available), respectively. Since the intervention participation effects could vary by examination, two-way interaction terms between intervention participation and examination were added to the models shown in Appendix A. A random effect at the student level was employed for the repeated measures structure. Three indicator variables were included to adjust for variability across the four examinations.

Cumulative reported participation models: The third approach accounted for the total number of study sessions attended by each student in the intervention groups. Students in the control group were excluded from Model 3. Since cumulative participation in study sessions was not complete until the end of the study, the outcome for this approach was the sum of the four examination scores (the cumulative examination score). This approach estimated the effect of intervention on cumulative examination score after adjusting for the number of study sessions in which the student reported participation. Since only one observation per student was required for this analysis, no repeated measures structure was necessary.

The session participation used in the second and third modeling approaches was based on self-report either at the time of completion of the self-evaluation problems after individual study sessions or during the post-study survey. Because self-report was not requested at the time of the first study session, the first session was not included as a separate time-point.

Each model was subsequently adjusted for baseline factors associated with performance which were identified from the pre-study survey (data not shown). Non-consenting students were not included in analyses of examination scores, according to the regulations of our investigational review board. However, completion of the pre-study survey was taken as tacit consent for that portion of the study among students who did not consent to join the whole study.

3. RESULTS

3.1. STUDY PARTICIPATION

A total of 376 students registered in the course; 265 (70%) of the students consented to enroll in the trial with 69, 100, and 96 randomized to the cooperative learning, internet learning, and control groups, respectively. Three students randomized to the internet learning group were excluded from the analysis due to early changes in student course plans, reducing the total number to 97.

The distributions of demographic and student characteristics for both randomized and non-enrolled students are shown in Table 1. As expected by randomization, all three groups were fairly comparable with respect to pre-study characteristics, with no statistically significant differences. In addition, few differences were found between students who consented to join the study and those who did not enroll but who still completed the pre-study survey. Approximately 49% of the non-enrolled students voluntarily completed the pre-study survey. The primary difference between these two groups was that non-enrolled students reported greater levels of concurrent employment.

Individual access to the study interventions was not tracked in either intervention group, although self-report of intervention participation was collected. In the cooperative learning group, the number of students present at each session was collected. In the internet learning group, the Superstats™ software was used to track overall access to the internet learning website over time (SiteCatalyst, 2002). Figure 2 compares the overall participation by session from these two methods. However, since the method of tracking participation differed by intervention group, comparisons in Figure 2 can only be made regarding overall patterns of participation, rather than the participation rate, because of differences in scale. Of the 69 students randomly assigned to the cooperative learning group, 45 (65%) attended the first session on September 13, 2001, two days after a national tragedy in the US.

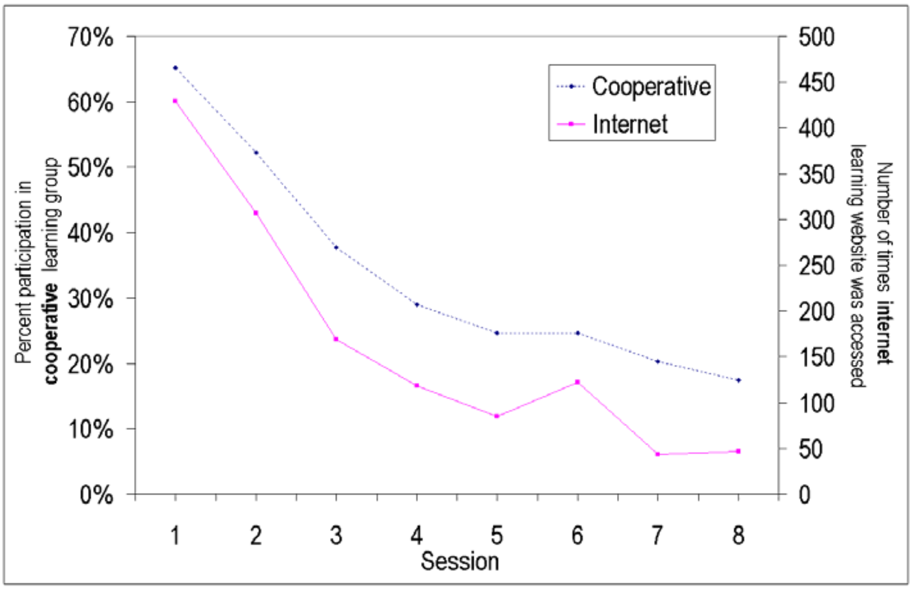

Figure 2. Participation in the cooperative learning and number of times the internet learning website was accessed, by study session

Table 1. Distributions of demographic and student characteristics for randomized and non-enrolled students

	Cooperative No. (%)	Internet No. (%)	Control No. (%)	p†	
Gender					
Male	20 (30.3)	27 (28.7)	30 (31.6)	0.91	19 (35.2)
Female	46 (69.7)	67 (71.3)	65 (68.4)		35 (64.8)
Age					
20-29	40 (58.0)	59 (60.8)	59 (61.5)	0.87	29 (53.7)
30-39	21 (30.4)	31 (32.0)	29 (30.2)		23 (42.6)
40-49	5 (7.3)	3 (3.1)	6 (6.3)		1 (1.9)
50+	3 (4.4)	4 (4.1)	2 (2.1)		1 (1.9)
Degree					
MPH	25 (37.9)	36 (38.3)	31 (32.6)	0.99	14 (25.9)
Other Masters	22(33.3)	32 (34.0)	36 (37.9)		19 (35.2)
Doctoral	12 (18.2)	17 (18.1)	17 (17.9)		12 (22.2)
Other	7 (10.6)	9 (9.6)	11 (11.6)		9 (16.7)
Credit Hours					
≤ 5	3 (4.4)	7 (7.2)	9 (9.4)	0.73	7 (13.0)
6-11	2 (2.9)	2 (2.1)	5 (5.2)		4 (7.4)
12-18	41 (59.4)	52 (53.6)	49 (51.0)		23 (42.6)
19+	23 (33.3)	36 (37.1)	33 (34.4)		20 (37.0)
English					
Native Language	42 (63.6)	52 (55.3)	60 (63.2)	0.45	32 (59.3)
Second Language	24 (36.4)	42 (44.7)	35 (36.8)		22 (40.7)
Employment					
10+ hours/week	24 (36.4)	30 (31.9)	41 (43.2)	0.28	44 (81.5)
<10 hours/week	42 (63.6)	64 (68.1)	54 (56.8)		10 (18.5)
	Mean (SD)	Mean (SD)	Mean (SD)	p‡	
Statistical Knowledge (Correct responses of 10)	4.28 (1.84)	4.32 (1.90)	3.71 (2.23)	0.078	4.44 (2.33)
Mathematical Skill (Correct responses of 5)	4.37 (1.22)	4.26 (1.23)	4.44 (1.02)	0.53	4.17 (1.19)
Desire for a Tutor (Likert scale: 0=definitely not needed to 4=definitely needed)	1.52 (0.96)	1.49 (0.89)	1.53 (1.11)	0.97	1.35 (1.07)
Total students	69	97	96		54

† Statistical significance for the difference between the randomized groups determined by Chi-square test.
‡ Statistical significance for the difference between the randomized groups determined by Analysis of Variance test.

3.2. STUDENT PREFORMANCE ON EXAMINATIONS

The overall mean cumulative examination score was 330.8 points (SD: 36.8 points). There was variability in mean score across the four course examinations. The overall mean (SD) scores were 89.0 (12.8) points; 81.3 (11.7) points; 82.8 (10.6) points; and 75.7 (14.1) points for the first through fourth examinations, respectively.

In a previous analysis variables from the pre-study survey were used to model cumulative examination score using forward stepwise regression incorporating two-way interaction terms. Younger age, greater mathematical aptitude (measured on a Likert scale from 0 to 5 based on a weighted scoring of the correct responses to questions 12 and 13 from Kemeny/Kurtz Math Series, 1992, p. 16) and statistical knowledge (measured on a 10 point scale adapted from Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000, and Wulff, Anderson, Brandenhoff, & Guttlet, 1987), working less than 10 hours per week, and student self-report of not needing a tutor (Likert scale of the reported need for a tutor; 0=definitely not, 1=probably not, 2=not sure, 3=probably, 4=definitely) were identified as pre-study factors associated with high performance. These five covariates were added in all subsequent models of intervention effect and performance.

Evaluating the Association of Intervention with Performance Based on Intent-to-Treat Models (1st Approach) No statistically significant differences in performance by study group were observed in the unadjusted intent-to-treat analyses. After adjusting for the five pre-study predictors of performance, estimated mean scores for students randomized to cooperative learning were 0.3 points below those of students randomized to control (95% CI: -3.4, 2.9); mean scores for students in the internet learning group were 0.01 points lower than students in the control group (95% CI: -2.8, 2.8).

Evaluating the Association of Intervention and Participation with Performance Based On Individual Reported Participation Models (2nd Approach) Table 2 shows results of the models of student performance on the four examination scores as a function of intervention and reported participation in sessions prior to the examinations. The results suggest increased performance in both intervention groups; however, statistically significant increases in performance were only observed at the time of the fourth examination. It should be noted that models for three or four consecutive study sessions could not be constructed for the first examination because only two study sessions had occurred by the time of that examination. All models in Table 2 included interaction terms of intervention effects and course examinations identified by Wald test results and were adjusted for the five pre-study predictors associated with performance.

Evaluating the Association of Intervention and Participation with Performance Based on Cumulative Reported Participation Models (3rd Approach) No statistically significant difference in performance (as measured by cumulative examination score) between the two intervention groups were observed after adjusting for the number of sessions the student reported attending (3rd Approach, see Table 3). However, performance increased with each additional study session in which the student participated. Each session was associated with a 2.1 point average increase (95% CI: 0.2, 3.9) in cumulative examination score in the adjusted model.

4. DISCUSSION

4.1. CONCLUSIONS

The goal of this study was to investigate whether the addition of active learning methods to a didactic introductory biostatistics course aided student understanding of key concepts, as measured by student performance on course examinations. The unadjusted intent-to-treat analysis revealed no statistically significant differences in performance across the three study groups (cooperative learning, internet learning, and control). This was likely attributable to low participation rates in the study interventions; by the third study session, 51% of the students in the two intervention groups had dropped out. From comments on the post-study survey, students in both intervention groups overwhelmingly cited lack of time as the predominant reason for nonparticipation. We also compared students who did not participate after the second study session with those who did complete later intervention sessions. The only difference found between participants and those who dropped out was that those completing later study sessions were enrolled in fewer credit hours.

In the presence of significant noncompliance, intent-to-treat analyses may not adequately reflect true differences between groups (Green, 2002). Accordingly, alternative analytic approaches were explored. The second modeling approach, using students' reported participation, suggested improved performance for participants as compared to nonparticipants and controls. The benefits of one study session were negligible. However, after four consecutive study sessions at the time of the fourth 100-point examination, cooperative learning participants scored an average of 5.3 points higher (95% CI: 0.4, 10.3), and internet learning participants scored an average of 8.1 points higher (95% CI: 3.0, 13.2), than nonparticipants or controls, after adjusting for the five pre-study factors associated with performance. The upper limit of the confidence interval reflects an improvement in understanding corresponding to perhaps two additional correct responses out of 20 examination questions. Under the 3rd modeling approach, each additional intervention session in which the student participated was associated with a 2.1 point increase in cumulative examination score (on a 400 point scale) (95% CI: 0.2, 3.9) in the adjusted model. When this effect is multiplied by the number of available intervention sessions, this increased performance may be substantial.

4.2. STRENGTHS AND LIMITATIONS

A limitation in the design of this study that could introduce bias was the requirement of extra work beyond the regular course material for the two intervention groups. One effect was decreased participation over time, which is associated with two potential biases; possibly students who continued to participate were more dedicated and thus more likely to work hard, or students who continued to participate did so because the intervention was more helpful to them than to those who dropped out. The effects of these potential biases may be most clearly observed in Table 2. By the time of the fourth examination, those who were still participating in the study interventions had likely participated in all four most recent study sessions; consequently, very little variation is observed in the increase in estimated performance from the models reflecting at least one study session versus the models reflecting at least four study sessions. Conversely, it is possible that students participating in the two intervention groups simply spent more time working with statistical concepts, and that additional time of any form would have led to the same improved performance.

Table 2. Linear models for students' subsequent examination scores by the number of prior study sessions attended (2nd Approach, interaction model)

	Intervention Group	Number of consecutive sessions attended	Change in Examination Score (95% CI)			
			Unadjusted Estimate		Adjusted Estimate*	
1st Examination	Cooperative Group vs. No Intervention	1 session	2.8	(-1.8, 7.3)	1.7	(-1.8, 5.2)
		2 sessions	4.8	(-0.3, 9.9)	2.6	(-1.6, 6.7)
	Internet Group vs. No Intervention	1 session	3.4	(-0.0, 6.9)	2.1	(-1.4, 5.5)
		2 sessions	3.1	(-0.4, 6.6)	1.7	(-1.7, 5.2)
2nd Examination	Cooperative Group vs. No Intervention	1 session	3.2	(-1.1, 7.4)	3.0	(-0.6, 6.6)
		2 sessions	3.4	(-1.2, 8.0)	3.1	(-0.8, 6.9)
		3 sessions	3.5	(-1.0, 8.0)	2.9	(-1.0, 6.8)
		4 sessions	4.1	(-0.4, 8.6)	2.9	(-1.2, 7.1)
	Internet Group vs. No Intervention	1 session	-0.7	(-4.0, 2.7)	-0.9	(-3.8, 2.0)
		2 sessions	-0.4	(-4.2, 3.3)	-0.9	(-4.3, 2.4)
		3 sessions	-0.2	(-4.0, 3.6)	-0.8	(-4.2, 2.6)
		4 sessions	-0.4	(-4.5, 3.7)	-0.9	(-4.7, 2.8)
3rd Examination	Cooperative Group vs. No Intervention	1 session	0.8	(-3.4, 5.0)	1.2	(-2.6, 5.0)
		2 sessions	1.3	(-3.2, 5.9)	1.7	(-2.4, 5.9)
		3 sessions	1.6	(-3.1, 6.2)	1.8	(-2.3, 5.9)
		4 sessions	1.6	(-3.0, 6.3)	1.7	(-2.4, 5.8)
	Internet Group vs. No Intervention	1 session	1.9	(-1.7, 5.4)	1.7	(-1.7, 5.2)
		2 sessions	2.3	(-1.4, 6.0)	2.1	(-1.5, 5.7)
		3 sessions	2.4	(-1.3, 6.3)	2.2	(-1.4, 5.8)
		4 sessions	2.6	(-1.3, 6.4)	2.1	(-1.6, 5.7)
4th Examination	Cooperative Group vs. No Intervention	1 session	3.6	(-2.7, 10.0)	4.5	(-0.4, 9.4)
		2 sessions	3.6	(-2.9, 10.0)	4.6	(-0.5, 9.6)
		3 sessions	4.2	(-2.3, 10.7)	*5.2*	*(0.1, 10.3)*
		4 sessions	4.3	(-2.2, 10.7)	*5.3*	*(0.4, 10.3)*
	Internet Group vs. No Intervention	1 session	*8.1*	*(2.9, 13.3)*	*7.1*	*(2.3, 11.9)*
		2 sessions	*8.8*	*(3.2, 14.4)*	*7.7*	*(2.6, 12.8)*
		3 sessions	*8.9*	*(3.3, 14.5)*	*7.9*	*(2.8, 12.9)*
		4 sessions	*9.0*	*(3.4, 14.6)*	*8.1*	*(3.0, 13.2)*

**Adjusted models include pre-study factors associated with performance.*

Table 3. Linear models for students' cumulative examination score by the number of prior study sessions attended (3rd Approach)

Comparison	Change in Cumulative Examination Score (95% CI)			
	Unadjusted Estimate		Adjusted Estimate*	
Cooperative Group vs. Internet Group	0.4	(-11.4, 12.2)	4.7	(-5.3, 14.7)
Each additional study session	1.9	(-0.3, 4.0)	*2.1*	*(0.2, 3.9)*

**Adjusted models include pre-study factors associated with performance.*

Another potential bias is the Hawthorne effect. Individuals who are aware that they are being studied may behave differently than they otherwise would (Franke & Kaul, 1978). The average examination score was 1.5 points higher among those randomized to the control group (95% CI: -2.0, 5.0), than among those who did not consent to enroll in the study. However, it was not possible to adjust this difference for the other performance predictors, since not all non-enrolled students completed the pre-study survey.

The key strength of this study design was the randomization of students to different study groups. Utilizing a longitudinal framework allowed comparison of the effects of the interventions on performance over time. The large initial sample size of the trial provided the power necessary for detecting differences in performance by intervention and participation in our analyses.

The inclusion and comparison of two different types of active learning (cooperative and internet) was another key component of this study. With the advent of distance education, technologically enhanced learning, such as interactive online applets, affords a new way to offer active learning within a distance-friendly format. In the internet learning group, no supervision was required and yet improved performance was observed that was comparable to that of the cooperative learning group, with far less intensive investment of instructor time. The website for the internet learning group required little resources other than providing an introductory interface and framework since publicly available applets were used. Development of new applets would have initial costs but require little maintenance and resources over time.

4.3. IMPLICATIONS FOR FUTURE INSTRUCTION AND RESEARCH

Conducting a randomized trial within the framework of a large class such as this was extremely challenging. Despite the large number of students who initially chose to join the study, overall participation was low. Given students' hectic schedules and demands on their time, participation in any optional educational research project will be limited. Increased participation is needed in future investigations. One option is to incorporate intervention materials as a required course component. A comparative study could be made of consecutive offerings of a course in which the second offering introduces required new material (such as active learning strategies) but otherwise the course remains the same. This approach was used by Smith (1998) to evaluate small group cooperative learning projects. Such a study design assumes no differences in student composition and requires the same assessment tools over time. However, the inclusion of such a comparison group is a critical part of evaluation of new statistical education methods.

Another consideration in the future evaluation of statistical education techniques is the specification of the amount of course content under evaluation. In this study, the intervention sessions covered a proportionately small amount of the course content and time relative to the other time requirements of the course. Consequently, only small changes in overall performance could be expected and their detection would require large sample sizes. Our choice of examination scores as the primary outcome variable resulted in high variability. Although we intended to utilize specific self-evaluation problems to evaluate the individual study sessions, these problems were not mandatory and thus were not completed by the majority of students. Future investigations warrant the incorporation of a required assessment tool that targets specific concepts emphasized through the intervention. One way to address this concern is the use of a hybrid course for the online portion of the intervention; however, ethical issues arise surrounding randomizing students to different types of courses.

These findings suggest that students may be aided by learning introductory biostatistics material via interactive activities especially if such activities are a required course component and offered throughout the term or semester. Our findings of an association of continued improvement in performance with completion of additional active learning sessions in the third approach is particularly encouraging. Cooperative learning activities and pertinent technological aids may both be helpful additions to on-site statistical education by either enhancing learning and/or reducing anxiety related to mathematical concepts. Future research and evaluation is needed to elucidate these relationships. In addition, research on online active learning methodologies is also required in the area of distance education. Continued development and evaluation of statistical teaching methodologies are critical and timely. Increasing numbers of public health professionals are seeking skills in quantitative methods and are faced with the challenge of mastering knowledge of appropriate statistical techniques and applications. The widespread availability of computer technology, both within and outside the classroom, provides an unparalleled environment for innovation in statistical education to maximize the potential for learning.

REFERENCES

Bradstreet, T. E. (1996). Teaching introductory statistics courses so that nonstatisticians experience statistical reasoning. *The American Statistician*, *50*(1), 69-78.

Franke, R. H., & Kaul, J. D. (1978). The Hawthorne experiments: First statistical interpretation. *American Sociological Review*, *43*(5), 623-643.

Garfield, J. (1993). Teaching statistics using small-group cooperative learning. *Journal of Statistics Education*, *1*(1). [Online: http://www.amstat.org/publications/jse]

Garfield, J. B. (1995a). Respondent: How should we be teaching statistics? *The American Statistician*, *49*(1), 18-20.

Garfield, J. (1995b). How students learn statistics. *International Statistical Review*, *63*(1), 25-34.

Gnanandesikan, M., Scheaffer, R. L., Watkins, A.E., & Witmer, J.A. (1997). An activity-based statistics course. *Journal of Statistics Education*, *5*(2). [Online: http://www.amstat.org/publications/jse]

Green, S. B. (2002). Design of randomized trials. *Epidemiologic Reviews*, *24*(1), 4-11.

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science*, *290*(5500), 2261-2262.

The Kemeny/Kurtz Math Series. (1992). *Algebra workbook*. USA: True Basic, Inc.

Kvam, P. H. (2000). The effect of active learning methods on student retention in engineering statistics. *The American Statistician*, *54*(2), 136-140.

Lovett, M. C., & Greenhouse, J. B. (2000). Applying cognitive theory to statistics instruction. *The American Statistician*, *54*(3), 196-206.

Magel, R. C. (1998). Using cooperative learning in a large introductory statistics class. *Journal of Statistics Education*, *6*(3). [Online: http://www.amstat.org/publications/jse]

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*. London: Chapman & Hall.

Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review*, *65*(2), 123-165.

Scheaffer, R. L., Gnanadesikan, M., Watkins, A., & Witmer, J. A. (1996). *Activity-based statistics: Instructor resources*. New York: Springer-Verlag New York, Inc.

Shaughnessy, J. M. (1977). Misconceptions of probability: An experiment with a small-group, activity-based, model building approach to introductory probability at the college level. *Educational Studies in Mathematics*, *8*, 295-316.

Simpson, J. M. (1995). Teaching statistics to non-specialists. *Statistics in Medicine*, *14*, 199-208.

SiteCatalyst (formerly *SuperStat*). (2002). Nashville, TN: Omniture, Inc. [Online: http://www.omniture.com/]

Smith, G. (1998). Learning statistics by doing statistics. *Journal of Statistics Education*, *6*(3). [Online: www.amstat.org/publications/jse/v6n3/smith.html]

The Stata Corporation. (2001). *Intercooled Stata version 7*, Houston Station, TX.

Utts, J., Sommer, B., Acredolo, C., Maher, M. W., & Matthews, H. R. (2003). A study comparing traditional and hybrid internet-based instruction in introductory statistics classes. *Journal of Statistics Education*, *11*(3). [Online: www.amstat.org/publications/jse/v11n3/utts.html]

Ward, B. (2004). The best of both worlds: A hybrid statistics course. *Journal of Statistics Education*, *12*(3). [Online: www.amstat.org/publications/jse/v12n3/ward.html]

Wulff, H. R., Anderson, B., Brandenhoff, P., & Guttlet, F. (1987). What do doctors know about statistics? *Statistics in Medicine*, *6*, 3-10.

FELICITY BOYD ENDERS, PHD, MPH
Division of Biostatistics
Division of Health Sciences Research
The Mayo Clinic
200 First Street, SW
Rochester, MN 55905, USA

APPENDIX A: VARIABLE DEFINITIONS AND MODELS USED FOR STATISTICAL ANALYSIS

Variables used in the models:

Y is the vector of examination scores for the four course examinations

(Exam I) through (Exam III) are indicator variables for the first three course examinations

(Coop) and (Internet) are indicator variables for randomization to the two study intervention groups

(Number) is the number of study sessions for which the student reported participation

Y_{cum} is the sum of all four course examination scores (cumulative examination score)

The following time-defined variables are each vectors of length seven, for which time is defined as t = 2, 3, 4, …8, representing study sessions two through eight.

$(Coop)_t$ and $(Internet)_t$ are the vectors of indicator variables for reported participation in the study intervention groups across study session two through eight (control group is the reference group)

$(Coop)_{t-1}$ and $(Internet)_{t-1}$ are the vectors of indicator variables for reported participation in the two study intervention groups at the time of the prior study session (control group is the reference group)

$(Coop)_{t-2}$ and $(Internet)_{t-2}$ are the vectors of indicator variables for reported participation in the two study intervention groups at the time of the second prior study session (with I=0 for $t \leq 2$) (control group is the reference group)

$(Coop)_{t-3}$ and $((Internet)_{t-3}$ are the vectors of indicator variables for reported participation in the two study intervention groups at the time of the third prior study session (with I=0 for $t \leq 3$) (control group is the reference group)

Y_t is the vector of examination scores for the first course examination after the current study session

$(Exam\ I)_t$ is the vector of indicator variables that the course examination following the current study session (at time t) was the first course examination [$I(Exam_t = Exam\ I)$]

$(Exam\ II)_t$ is the vector of indicator variables that the course examination following the current study session (at time t) was the second course examination [$I(Exam_t = Exam\ II)$]

$(Exam\ III)_t$ is the vector of indicator variables that the course examination following the current study session (at time t) was the third course examination [$I(Exam_t = Exam\ III)$]

1: Intent-to-treat model

$$\text{Model 1. } E[Y] = \alpha_0 + \beta_1(\text{Coop}) + \beta_2(\text{Internet}) + \gamma_1(\text{Exam I}) + \gamma_2(\text{Exam II}) + \gamma_3(\text{Exam III}) + \varepsilon_{\text{intra-student}} + \varepsilon_{\text{inter-student}}$$

2. Individual reported participation models

$$\text{Model 2a. } E[Y_t] = \alpha_0 + \beta_1(\text{Coop})_t + \beta_2(\text{Internet})_t + \gamma_1(\text{Exam I})_t + \gamma_2(\text{Exam II})_t + \gamma_3(\text{Exam III})_t + \varepsilon_{\text{intra-student}} + \varepsilon_{\text{inter-student}}$$

Model 2b. $E[Y_t] = \alpha_0 + \beta_1(\text{Coop})_t + \beta_2(\text{Internet})_t + \beta_3(\text{Coop})_{t-1} + \beta_4(\text{Internet})_{t-1}$
$+ \gamma_1(\text{Exam I})_t + \gamma_2(\text{Exam II})_t + \gamma_3(\text{Exam III})_t$
$+ \varepsilon_{\text{intra-student}} + \varepsilon_{\text{inter-student}}$

Model2c. $E[Y_t] = \alpha_0 + \beta_1(\text{Coop})_t + \beta_2(\text{Internet})_t + \beta_3(\text{Coop})_{t-1} + \beta_4(\text{Internet})_{t-1}$
$+ \beta_5(\text{Coop})_{t-2} + \beta_6(\text{Internet})_{t-2} + \gamma_1(\text{ExamI})_t + \gamma_2(\text{ExamII})_t + \gamma_3(\text{ExamIII})_t$
$+ \varepsilon_{\text{intra-student}} + \varepsilon_{\text{inter-student}}$

Model 2d. $E[Y_t] = \alpha_0 + \beta_1(\text{Coop})_t + \beta_2(\text{Internet})_t + \beta_3(\text{Coop})_{t-1} + \beta_4(\text{Internet})_{t-1}$
$+ \beta_5(\text{Coop})_{t-2} + \beta_6(\text{Internet})_{t-2} + \beta_7(\text{Coop})_{t-3} + \beta_8(\text{Internet})_{t-3}$
$+ \gamma_1(\text{Exam I})_t + \gamma_2(\text{Exam II})_t + \gamma_3(\text{Exam III})_t$
$+ \varepsilon_{\text{intra-student}} + \varepsilon_{\text{inter-student}}$

3. Cumulative reported participation models

$E[Y_{\text{Cum}}] = \alpha_0 + \beta_1(\text{Coop}) + \beta_2(\text{Number})$
$+ \varepsilon_{\text{inter-student}}$