COMPARISON OF MULTIMEDIA EDUCATIONAL MATERIALS USED IN AN INTRODUCTORY STATISTICAL METHODS COURSE

J. Richard Alldredge and Nicholas A. Som
Washington State University
USA

We desired to improve student learning in our introductory, algebra-based statistical methods course. Marketing claims as well as anecdotal evidence suggested that electronic forms of educational material improve student learning. Some recent empirical evidence presented in the statistical literature uses both qualitative and quantitative data to evaluate computer-based learning aids. In order to contribute to evaluation of educational technology we designed an experiment to evaluate the use of ActivStats Multimedia Educational Software (on CD) and CyberStats Introduction to Statistics (on the web). Specifically, we assessed how the use of these two forms of educational material in the statistics laboratory portion of our course impacted student learning and attitudes. Implications of these results are discussed as well as lessons learned for designing future experiments.

INTRODUCTION

The assessment of new instructional methods, theories, and tools to improve student learning are prevalent ideas in current statistical education literature. The objective of this work is to further that assessment. Because instructors of statistics are looking for ways to help their students better learn statistics (Smith, 1998), it seemed appropriate to analyze some of the tools that are being used to improve student learning. According to Jolliffe (1998), "Thus educational research activities in computing science and in learning technology provide another outlet for research in statistical education." With this in mind, an experiment was designed to compare two computer based educational statistical packages. Our hope was that this classroom research project would yield valuable insights into the teaching and learning of statistics to benefit the education community as suggested by delMas, Garfield, and Chance (1999).

The purpose of the experiment was to determine if two educational statistical packages differed in their ability to improve statistical learning and attitude toward learning statistics. The experiment was implemented in an introductory algebra-based statistical methods course during the Fall 2000 semester. This course was designed to fit the needs of many students. It satisfies a general education requirement for mathematics proficiency at Washington State University. In addition, this course satisfied the requirement of many departments that students take one course in statistics. The students came from broad backgrounds in previous mathematical and statistical knowledge, current academic interests, age, and demographics. The course consisted of two components. The first component of the class was three semester hours of lecture instruction per week, and the second was a two-hour weekly laboratory session. There were two lecture sections of the class. One section was divided into three laboratory sections and the other larger lecture class was divided into six lecture sections. Each laboratory section was assigned one of the two computer-based programs to be used in the laboratory for the entire semester.

THE EXPERIMENT

Many issues bearing pedagogical importance were considered in the planning of this experiment. In order to minimize instructor influence on overall differences among the attitudes and performances of the students, a single instructor volunteered to teach both lecture sections of the course during the Fall 2000 semester. The same textbook was used for both lecture sections. Each of the three teaching assistants that ran the laboratory sessions was assigned an equal number of sections from each instructional package. There were several instructional packages to choose from for this experiment. We agree with Lee (1998) that introductory statistics should be taught using real world data, student activities, and computer technology. We therefore selected CyberStats and ActivStats as the two products to compare. According to product information accompanying the software, some of the ActivStats claims are that students will experience real world examples, learn key statistics concepts through specially designed simulations, and practice with interactive experiments. On the CyberStats web page the following principles are listed: learning by activity and discovery, real data in real-world settings, and a stress on conceptual understanding. Clearly, both packages make similar claims, and there are other similarities. Each package also contains its own version of a computational statistics program that both interfaces with the topical lessons, and is available for use independently of the instructional activities. Both also offer program customization, such as ordering topics to match texts when used as complimentary resources, as was the case with this experiment. CyberStats is a world-wide-web based program. Students pay a fee for a password that gives them access to the material for the duration of the academic term. ActivStats is purchased on a CD-Rom, and is the students to keep at the end of the semester, just as any textbook. The cost for each program was comparable, and each student was required to obtain their own copy or access password for their respective package.

In order to qualify and quantify the differences between the two packages, numerous assessment tools were used. At the first laboratory session, the students filled out a questionnaire with 39 questions addressing issues of quantitative, verbal, and academic confidence. The questionnaire also addressed computer proficiency and their feelings considering applications of statistics and academics to their future. Possible responses were coded as a 5 for strongly agree, 4 for agree, 3 for somewhat agree, 2 for disagree, and 1 for strongly disagree. A similar questionnaire was given during the final laboratory session. In addition to covering some of the same issues, the post-class questionnaire also included questions regarding their experiences with the computer-based program to which they were assigned. Responses were coded the same as for the initial survey. A student's pre-course survey score was subtracted from their post-course survey score, creating a difference score for the questions common to both surveys. We added 5 to this score to eliminate negative numbers so the resulting difference scores could range from 1 to 9. The difference score was used to assess change in attitude and opinion during the semester.

Student assessment scores used to assess learning included two mid-semester exams based on topics covered in both the lecture and laboratory portions of the class. Each exam focused on subject matter presented up to the point that the exam was given. The second exam, though focusing mainly on topics covered after the first exam, still required the retention of material from the beginning of the course and applied to the new topics, such as mean and standard deviation. A cumulative final exam was administered at the end of the semester and these scores were also used to compare the two instructional packages. For all three exams, each student was allowed to create a single help sheet for their own aid, and all necessary tables were provided. An additional comparison between packages used total course points. This total included the mid-semester exam scores, the final exam scores, as well as scores compiled from in-class and laboratory activities, lecture and laboratory homework assignments, and two class projects. Each project involved the application of statistical topics covered during the semester. The projects, although containing statistical analysis, were largely written works and graded for pertinent statistical content, quality of writing, and for meeting the requirements laid out in the project description.

The treatments for this experiment, ActivStats and CyberStats, were assigned to laboratory sections. That is, all students in a laboratory section used the same package. For administrative convenience ActivStats was used by the three laboratory sections associated with the smaller lecture class and CyberStats was used by the six laboratory sections associated with the larger lecture class. Because the treatments were applied to laboratory sections, rather than to individual students, the laboratory section was considered the experimental unit for comparing instructional packages. Characteristics of students entering the course, such as quantitative and verbal skills, varied widely. Analysis of covariance was used to adjust for pre-course student differences and allow a better comparison of the instructional package treatment. Covariates measured on each student included a preliminary quantitative quiz completed during the first lecture class meeting, SAT verbal score, SAT math score, SAT total score, and a university admission index number.

ANALYSES AND RESULTS

One question of interest was: Did the changes in student attitudes and opinions differ depending on the instructional package they used? For this analysis, the differences in responses to questions common to both the pre-course and post-course survey were analyzed. A two-level model was analyzed with analysis of covariance. An instructional package treatment was applied to each laboratory section at the first level. In the second level, covariates were measured on individual students within laboratory sections. The results of the analysis revealed that students in laboratories using ActivStats gained significantly more confidence in writing ability and ability to do well on exams than those in laboratories using CyberStats. The treatment means, p-values, and significant covariates associated with these results can be found in Table 1.

Table 1
Means, P-values, and Covariates for Differences in Post-course and Pre-course Questionnaire Responses

Difference Variable*	ActivStats Mean	CyberStats Mean	P-value	Covariate(Significance)
Z1	4.37	3.91	.0754	None
Z2	4.37	3.88	.0486	SAT Verbal (.0066)
Z2	4.37	3.86	.0409	SAT Total (.0186)

*Z1: This semester, I have gained confidence in my ability to do well on exams.
Z2: This semester, I have gained confidence in my ability to write well.

The second analysis concerning the surveys focused solely on the post-course questionnaire. In order to reduce the dimensionality of the survey and identify the underlying patterns of variation in the data set, a multivariate principal component analysis was conducted. Principal components analysis can be used to assess the actual dimensionality, and produce a smaller number of uncorrelated variables to be used in analyses (Johnson, 1998). The pattern and size of the coefficients for the principal components allowed labeling of six groups of survey questions that summarized the variation in the data. These results were used to create new linear combinations of response variables that were in turn analyzed with covariates. One of the new linear combinations of survey questions was associated with feelings of confidence by students in their ability to write well, use math skills, and take exams. The results of the analysis revealed that students in laboratories using ActivStats gained significantly more confidence than those where CyberStats was used. Another of the groupings identified by the principal component analysis involved questions about the applicability of statistics to their future careers and its need for future success. An analysis of covariance indicated that using ActivStats made a significant difference in opinions on the applicability of statistics and the need of statistics for future success. The treatment means, p-values, and significant covariates associated with these results can be found in Table 2.

Table 2
Means, P-values, and Covariates for Principal Components

Principal Component	ActivStats Mean	CyberStats Mean	P-value	Covariate (Significance)
Confidence	7.97	7.16	.0390	SAT Math (.0006)
Confidence	8.01	7.13	.0288	SAT Total (.0004)
Confidence	8.05	7.10	.0346	SAT Verbal (.0022)
Confidence	7.99	7.28	.0542	Scores (.0370)
Future Success, Applicability	4.99	4.49	.0886	SAT Math (.0024)
Future Success, Applicability	5.03	4.47	.0635	SAT Total (.0009)
Future Success, Applicability	5.06	4.45	.0474	SAT Verbal (.0023)

The next analysis dealt with testing how the exam scores and total course points may have been affected by each instructional package. Again covariates were used, and the results suggest significantly higher scores on the first exam, second exam, final exam, and total course points for those students in laboratories that used ActivStats compared to students in the CyberStats laboratories. The treatment means, p-values, and significant covariates associated with these results can be found in Table 3.

Table 3
Means, P-values, and Covariates for Exam Scores and Total Course Points

Assessment	ActivStats Mean	CyberStats Mean	P-value	Covariate (Significance)
Exam 1	82.32	75.16	.0485	SAT Verbal (<.0001)
Exam 1	81.70	75.50	.0459	SAT Total (<.0001)
Exam 2	78.48	72.52	.0739	SAT Verbal (<.0001)
Exam 2	77.90	72.90	.0993	SAT Total (<.0001)
Final Exam	80.50	76.40	.0945	SAT Total (<.0001)
Total Course Points	854.59	816.65	.0466	SAT Verbal (.0004)
Total Course Points	852.20	818.40	.0766	SAT Total (<.0001)

CONCLUSION

In this experimental setting it appears that ActivStats was a better instructional tool. Not only did ActivStats seem to improve confidence in writing, taking exams, and confidence in general, but ActivStats seems to have led to higher exam scores and total course points than CyberStats. It is not clear to the authors why these differences occurred. The philosophy behind both packages is similar. Both emphasize activity based learning, real world situations, and conceptual understanding. Both packages were used in the weekly laboratory class and students were encouraged to use their respective package outside the laboratory as well. ActivStats uses cartoon style characters and noises to reward correct answers to activities whereas CyberStats provides written examples of solutions for students to self-check their work. ActivStats could be described as warmer and more friendly than CyberStats, or alternatively, as being designed for a less mature audience. It is not obvious that the targeted maturity level of the package can explain the significant differences that were observed. Another possible explanation for the differences could be that the students in the two lecture sections were different in ways that could not be adjusted for by the observed covariates. Also, the class that received the CyberStats treatment met for 75 minutes twice per week in an auditorium, and contained about twice as many students as the class receiving the ActivStats treatment. The lecture class receiving ActivStats met for 50 minutes three times per week in a smaller, more traditional classroom setting. As a consequence, differences in the physical environment for lectures were completely confounded with the instructional package treatments.

In the process of understanding our analytic results, several other possible analyses came to mind. Recall that questionnaire responses by students showed there was a significant difference in their confidence in their ability to write well. Also SAT verbal was identified as a significant covariate. To further explore this connection to verbal skills and writing we plan to analyze the scores on the two written projects. Another interest of the authors is to determine if the effect of the instructional package is different depending on student gender. We also desire to examine the effect the different teaching assistants may have had on the comparison of the instructional packages. Finally, a detailed examination of written student comments concerning the instructional package they used, as well as comments about the course in general, is planned. These comments may provide insights into differences in the instructional packages.

Educators looking for ways to improve student learning are turning to technology to capture the interest and maximize the learning potential of their students. Major concerns for statistics educators are to ensure that students understand statistical concepts and understand how to apply them to real-world situations (Garfield, 1995). The objective of this study was to design an experiment to compare two statistical instructional packages that make similar claims about improving student learning. The question of interest is not whether to use technology as an instructional aid, but rather, which technology to use, how much technology to use, and what to expect from technology in the classroom. As statisticians and educators, it makes sense to use statistical tools to evaluate new pedagogical developments so that refinement and further discoveries occur more rapidly and effectively. The results of this study indicate that differences in student attitudes and class performance may be detected when different instructional aids are used. Investigation continues into the causes for these differences. Future experiments could focus on how each treatment may have affected the learning of specific statistical topics covered in the class and aid in understanding the effects of each treatment. Additional experiments are being designed to provide more information, new insights, and new direction for future research.

ACKNOWLEDGEMENTS
Hayden Stewart, Korash Hernandez, and the Fall 2000 Stat/Math 212 students.

REFERENCES
delMas, R.C., Garfield, J., & Chance, B.L. (1999). A model of classroom research in action: developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education* (electronic journal), *7*(3).

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

Johnson, D.E. (1998). *Applied multivariate methods for data analysts.* Pacific Grove, CA: Brooks/Cole Pub.

Jolliffe, F. (1998). What is research in statistics education? In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee and W. Wong (Eds.), *Proceedings of the 5th International Conference on Teaching Statistics,* Vol. II, (pp. 801-806). Voorburg, The Netherlands: ISI.

Lee, C. (1998). An assessment of the PACE strategy for an introductory statistics course. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee and W. Wong (Eds.), *Proceedings of the 5th International Conference on Teaching Statistics,* Vol. III, (pp. 1215-1222). Voorburg, The Netherlands: ISI.

Smith, G. (1998). Learning statistics by doing statistics. *Journal of Statistics Education* (electronic journal), *6*(2).