Comparing the Effectiveness of Traditional and Active Learning Methods in Business Statistics: Convergence to the Mean

David Weltman
Mary Whiteside
The University of Texas at Arlington

Journal of Statistics Education Volume 18, Number 1 (2010),
www.amstat.org/publications/jse/v18n1/weltman.pdf

Key Words: Active learning; Teaching statistics; Student grade point average; Linear mixed models

Abstract

This research shows that active learning is not universally effective and, in fact, may inhibit learning for certain types of students. The results of this study show that as increased levels of active learning are utilized, student test scores decrease for those with a high grade point average. In contrast, test scores increase as active learning is introduced for students in the lower level grade point average group. Every student involved in the experiment is taught three topics, each one by a different teaching method. Students take a test following each learning session to assess comprehension. The experiment involves more than 300 business statistics students in seven class sections. Method topic combinations are randomly assigned to class sections so that each student in every class section is exposed to all three experimental teaching methods. The effect of method on student score is not consistent across grade point average. Performance of students at three different grade point average levels tended to converge around the overall mean when learning was obtained in an active learning environment. The effects of the teaching method on score do not depend on other student characteristics analyzed (i.e. gender, learning style, or ethnicity). A linear mixed model is used in the analysis of results.

1. Introduction

This research quantitatively measures student learning in each of three varied learning environments: a traditional lecture format, a hybrid format, and a fully active learning workshop. Each student in a core quantitative business school course is tested in each learning environment. It is assumed that students would learn the most about a topic in the fully active environment, but this research shows that the effectiveness of a method depends on an important student characteristic. We cannot simply assume that one method is more effective than another. In this study, the effectiveness of a teaching method depends on the student's cumulative grade point average level. This research demonstrates that active learning tends to equalize students of all levels. There is a convergence to an overall mean. Lower-level-student performance rises, while performance for the higher-level-student group significantly declines as more active learning elements are introduced.

2. Background

New learning tools and techniques, such as active or experiential learning, which have the potential to enhance an educational environment are of particular interest to university researchers (Lee 2007; Barak, Lipson and Lerman 2006; Hansen 2006; Raelin & Coghlan 2006). Although active learning as a concept dates back centuries, in modern times it was first described in detail by the English scholar R.W. Revans (1971) who further developed the concept over the following two decades. Briefly, Revans refers to active learning as reflection on experience and states that learning is achieved through focusing on problems in a social context (Revans, 1983), i.e. managers learning from each other and enhancing learning through interaction and shared experiences. More recently, Bonwell and Eison (1991) define active learning as "instructional activities involving students in doing things and thinking about what they are doing." The concept of active learning continues to evolve over time.

The traditional classroom lecture has been a dominant teaching method in business schools for decades (Alsop, 2006; Becker, 1997; Brown & Guilding, 1993). Current assessments of this technique show potential for improvement to this long-standing tradition (Bonwell, 1997). In fact, a number of business schools employ participant-centered and case-based learning. These methods are especially popular in graduate programs (i.e. MBA curriculum). Increasing competition among business schools, rising student expectations about teaching, and students seeking an active, high-impact learning experience in the classroom all contribute to this shift (Auster & Wylie, 2006). Furthermore, many scholars also note that business students are demanding more engaging learning experiences (O'Brien & Hart, 1999; Page & Mukherjee, 2000; Schneider, 2001).

Several studies (Lee 2007; Raelin & Coghlan, 2006; Sarason & Banbury, 2004; Sutherland & Bonwell, 1996; Ueltschy, 2001; Umble & Umble, 2004) have demonstrated both quantitative and anecdotal evidence regarding the effectiveness of active learning techniques. This research further develops understanding of active learning effects by empirically analyzing data obtained by conducting a semester-long experiment in a quantitative business school course (undergraduate business statistics). Subject characteristics such as gender, ethnicity, learning style, and grade point average are used to determine which characteristics are important in estimating how well a student performs under a particular teaching method. All students receive each of the three teaching methods and the treatments are randomized to class section. Thus, some students receive instruction in a topic under a certain teaching method; whereas, other students receive instruction in the same topic under a different teaching method.

Students appear to favor new methods of learning over the more traditional methods although a significant amount of the business research in active learning is anecdotal in nature. Our research suggests that in a quantitative undergraduate business course, active learning methods may not be effective at all and, in fact, may degrade the learning of students with higher overall GPAs. This research provides specific empirical evidence regarding the effectiveness of such new experiences.

3. Experimental Approach

Three core topics in the business statistics course are involved in the experiment: the binomial distribution, sampling distributions, and the calculation of p-values in hypothesis testing. These topics cover a fairly wide range in level of difficulty. In general, we sought topics that were complex enough to challenge all types of students so that we could obtain a range of rich and varied outcomes.

This research uses three teaching methods: a “traditional” lecture, a hybrid format, and a fully-active workshop. A teaching method is administered in a 60-minute session where a particular experimental topic is covered. The traditional lecture teaching method is likely the most widely used technique in business schools today. In this teaching method, students sit and listen to a lecture that has been structured and prepared by the instructor. In the current study, lecture slides prepared by the researcher were consistent across instructor and topic combinations.

Sutherland and Bonwell (1996) suggest that incorporating short experiential learning activities into a traditional lecture may be an effective way to gain many of the benefits of active learning with a minimum amount of disruption to the familiar lecture. This environment is referred to as the hybrid format method. Effective strategies for this technique include: the pause procedure, short writes, and think-pair-share. These authors argue that after about 15 minutes of lecture, students' ability to assimilate material rapidly declines. This research implements the hybrid method through the use of the traditional method punctuated with several breaks for students to collaborate on questions posed by the instructor. These lecture pauses focus on applications of and computations based on statistical methods. For example, after about 15 minutes of lecture regarding the binomial distribution, students are asked to come up with example situations where application of the binomial distribution would be appropriate. After another approximately 15 minutes of lecture, students are asked to calculate simple binomial probabilities that are discussed.

A fully-active workshop has the highest level of student involvement of the three teaching method formats. In the workshop, students work in small teams of two or three utilizing documentation that has been developed by the researcher. Here, the instructor works more as a "consultant" than a lecturer. Students are responsible for their own learning but have an expert available to answer questions and provide guidance concerning a particular topic. For example, in the sampling distribution workshop, students work in pairs using software that interactively displays sampling distributions from different population distributions for selected sample sizes.

A 20-minute multiple choice quiz follows an experimental session. Subject performance is measured by the percent of questions answered correctly. Questions are designed to assess a fairly wide variety of skills obtained. Questions are designed to test relatively simple skills, such as the ability to recall and define, as well as much more complex skills, such as the ability to compare, apply, and employ techniques appropriately.

A linear mixed model is chosen for analysis since study factors have both fixed and random effects. Prior to commencement of the experiment, it was assumed that the extraneous variation associated with topic and class section would be important. Our consolidated research model is

$$Y_{i(j)mt} = \beta_0 + \beta_1 M_1 + \beta_2 M_2 + \beta_4 A + \beta_7 AM_1 + \beta_8 AM_2 + S_{i(j)} + T_t + \varepsilon_{i(j)mt},$$

where $Y_{i(j)mt}$ is the subject's test score measured as percent correct for student i nested in section j, for method m, and for topic t. M_1 and M_2 are method indicator variables and A is the continuous covariate for grade point average. Each subject is tested a total of three times, and thus, there are three repeated Y measurements for each student. The three test scores for each subject are possibly correlated as the stronger students may have higher scores than weaker students. A high score in one method-topic combination might be associated with high scores in the other two method-topic combinations. The proposed mixed model accounts for these possible correlations.

In addition to cumulative grade point average, this research considers three other student characteristics: gender, learning style, and ethnicity. Students took a short online test to determine their dominant learning style: visual, aural, read/write, or kinesthetic. We are interested in determining whether or not the effect of method on score depends on the characteristic, i.e. interaction. These effects are tested with the base research model

$$Y_{i(j)mt}(\underline{x}) = \underline{x}\underline{\beta} + S_{i(j)} + T_t + \varepsilon_{i(j)mt},$$

where $\underline{x}$ in the first model contains gender characteristic main effects as well as the interaction of gender and method. Similarly, the second model contains learning style main effects as well as the interaction of learning style and method. The third model contains the ethnicity characteristic and its interaction with method. All three models include the continuous covariate GPA and its interaction with method. Gender, learning style, and ethnicity as well as associated interactions are treated as fixed effects. The results indicate the interactions of method with gender, learning style, or ethnicity are not significant, and thus, we have not included them in our consolidated research model. However, the interaction of the method with student grade point average is significant. We seek to understand how differences in the effects of the three teaching methods on learning depend on the subject's grade point average. Consequently, method by GPA interactions are retained in our consolidated model. We test whether or not there are significant effects in student learning because of teaching method, while controlling for any potential random effects of topic and student nested in class section.

To calculate reasonable confidence interval estimates for significant differences found between different grade-point-average group method combinations, we use the Bonferroni procedure. For mixed models such as the one described in the present study, Satterthwaite approximations for degrees of freedom are recommended (Dean & Voss, 1999; Verbeke & Molenberghs, 1997; West, Welch, & Galecki, 2007).

To test the significance of the random effects terms, a likelihood ratio test is employed. To test these hypotheses, we compare the –2 log-likelihood value for a reference model to a –2 log-likelihood value for a model which omits the random class section or random topic effect. The asymptotic null distribution of the test statistic is a mixture of χ^2 distributions, with 0 and 1 degrees of freedom, and equal weights of .5 (Verbeke & Molenberghs, 1997; West, Welch, & Galecki, 2007). Restricted or residual maximum likelihood estimation (REML) introduced by Patterson and Thompson (1971) is the method used in estimating variance components and testing the random effects since our study involves an unbalanced design. REML is preferred to maximum likelihood estimation for testing random effects because it produces unbiased estimates of covariance parameters by taking into account the loss of degrees of freedom that result from estimating the fixed effects in β (West, Welch, & Galecki, 2007). If the difference in the two models (full versus reduced) is represented by the symbol d, then the p-value for the likelihood ratio test statistic is given by

$$p-value = 0.5 * P(\chi_0^2 > d) + 0.5 * P(\chi_1^2 > d).$$

Using residual maximum likelihood estimation, our results indicate retention of the random effect terms.

The underlying strategy for question development and student assessment in this experiment is Bloom's taxonomy of cognitive domain (Bloom 1956). Questions are intended to assess a fairly wide variety of skills obtained. Questions are variously designed to test relatively simple skills, such as the ability to recall and define, as well as more complex skills, such as the ability to compare, apply, and utilize techniques appropriately. Overall, the majority of questions fall into the Bloom taxonomy of comprehension, application, and analysis, with a few questions assessing higher and lower levels of learning. Questions are constructed to assess student grasp of mechanical and conceptual issues. The Appendix shows sample assessment questions and our categorization of where they fall in the Bloom taxonomy. Although this research does not directly address whether or not a particular method favors higher or lower levels of learning in the taxonomy, a future, similar experiment may be performed with additional assessment questions to explore this interesting topic.

Every effort is made to select topic areas that cover a fairly wide range in level of difficulty (mechanical, applied, and conceptual). In general, we seek topics that are complex enough to challenge all types of undergraduate students so that we can obtain a range of outcomes. In the binomial distribution content, students are taught the basic ideas that underlie the distribution, its appropriate application, related assumptions, and basic calculations of binomial probabilities for various scenarios. The mean, variance, and causes of skewing are also covered in this topic. In the second topic, students learn the basic ideas of sampling distributions. The central limit theorem and the concept of standard error are covered. In the p-value approach to hypothesis testing, students are exposed to the idea of a test statistic under conditions where the population standard deviation is both known and unknown. Various hypothesis-testing scenarios are covered and techniques for calculating p-values are explained by using tables (the standard normal and student t) and Microsoft Excel.

4. Analysis of Results

4.1 Main Teaching Method Effects

Overall, the traditional method produces the highest test scores, followed by the hybrid method, then collaborative workshop. There is a significant difference in the effects of the different teaching methods, and these effects differ depending on student grade point average.

The main effects are examined by conducting two tests: a likelihood ratio and Type III F-test. Some researchers (Fai & Cornelius, 1996; Verbeke & Molenberghs, 1997; West, Welch, & Galecki, 2007) suggest the appropriateness of both tests for analyzing significance of fixed main effects. Likelihood ratio tests are used in maximum likelihood estimation; Type III F-tests are used in restricted maximum likelihood estimation. The results of all the tests are consistent and suggest significant teaching method effects. Confidence intervals for possible ranges of the teaching method effects are not presented here since it will be shown that the effect of the teaching method depends on student grade point average.

4.2 Interaction of Grade Point Average by Method

Method, cumulative GPA, and the interaction of method and GPA are all significant after controlling for the random effects of subject nested in class section and topic. Parameter estimates are used to determine point estimate scores for three different grade point averages, 1.75, 2.75, and 3.75, with each of the three teaching methods.

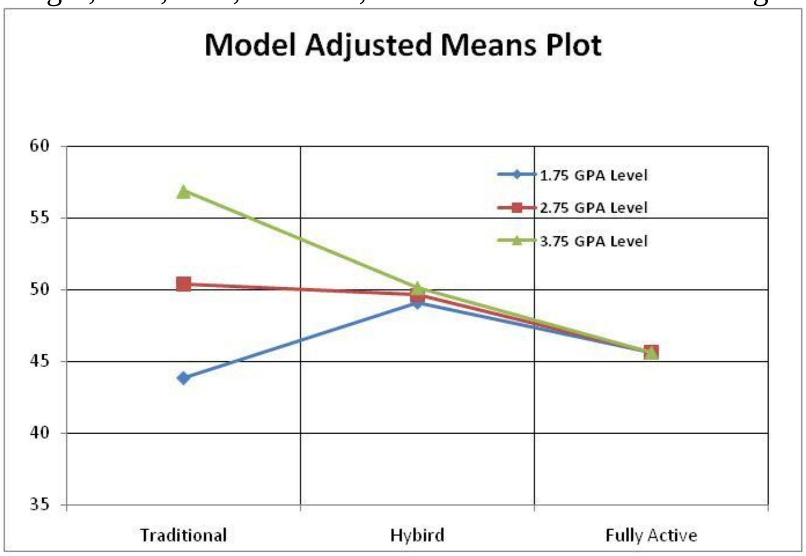

Figure 1. Convergence to the mean.

Figure 1 is a plot of the model-adjusted means. After adjusting for class section, topic, and individual variation, students with a high grade point average scored lower when exposed to the active learning methods versus the traditional teaching method.

Conversely, students in the low-grade-point-average category scored lowest when receiving the traditional method of teaching and higher in the active learning environments. Students in the mid-range grade point average group tended to score lower in the active learning environments (after adjusting for the random effects of section, topic, and student), but this trend was not as pronounced as in the high GPA group.

To study further the method by GPA interaction effect, six simultaneous confidence intervals are derived. The Bonferroni multiple comparison procedure is chosen for this part of the analysis since there are relatively few (six) comparisons of interest.

Table 1. Confidence Intervals of Mean Differences in Methods for Three GPA Categories with a Family Confidence Coefficient of .90

Test	Estimate	B	SE	LL	UL
3.75 GPA M1 vs M2	6.749	2.394	2.6856	0.319674	13.17833
3.75 GPA M1 vs M3	11.267	2.394	2.6724	4.869274	17.66473
2.75 GPA M1 vs M2	0.746	2.394	1.822	-3.61587	5.107868
2.75 GPA M1 vs M3	4.7305	2.394	1.8293	0.351156	9.109844
1.75 GPA M1 vs M2	-5.257	2.394	3.0231	-12.4943	1.980301
1.75 GPA M1 vs M3	-1.806	2.394	2.9665	-8.9078	5.295801

For students in the high-GPA group, mean scores under method 1 (the traditional teaching method) are higher than those scores under method 3 (the fully active teaching method) by somewhere between 4.8 and 17.7 points, after controlling for class section and topic. Additionally, for students in the high-GPA group, mean scores under method 1 are higher than those scores under method 2 by up to 13 points. And, for students in the mid-level-GPA group, mean scores under method 1 are higher than those scores under method 3 (the fully active teaching method) by somewhere up to 9 points, after controlling for class section and topic. No significant effects are found across methods for the low GPA group. However, it can be said that students in the low GPA group seem to perform better in the active learning methods versus the traditional method of teaching.

5. Conclusions and Implications in Research

There appears to be a phenomenon of convergence. As higher-GPA-level students are exposed to more classroom use of active learning methods, their level of learning drops to around the overall mean. The opposite is true for the low-level-GPA students whose scores improved, though not significantly, with higher levels of active learning.

Students with a higher cumulative grade point average scored higher on the tests when receiving the traditional method of teaching versus one of the active methods. Students with low cumulative grade point average scored worse under the traditional method. Our results show dependence on the level of a student's cumulative grade point average and are interpreted in this context.

The results of this research suggest that cumulative grade point average (or a surrogate) is an important research variable and should be included in models developed to analyze the effects of active learning. In this study, grade point average interacts with or moderates the relationship between teaching method and student learning (i.e. quiz score). This effect occurs in such a way that there is a convergence to a number near the overall mean in learning as students are exposed to more intense levels of active learning. This convergence finding is surprising, and one may speculate as to reasons for its occurrence. It is possible that students with a high grade point average achieve a deeper level of learning when experiencing exposure to the maximum amount of instructor expertise and direction. This result may be especially true in a quantitative class, such as the one in this study. In this study, the traditional method, as opposed to the active methods, offered such an experience.

Another explanation of the convergence phenomenon is that students with a high grade point average have, for the most part, attended courses where the dominant teaching paradigm is the traditional method. These students are not accustomed to learning outside of the traditional method, and thus their scores drop in alternative methods. One could also argue that the high-GPA-level students have a high grade point average because these students learn best with traditional learning and this teaching method is currently practiced in most business school undergraduate curricula. Thus, the reason for the drop in learning for the high-level-GPA students may be twofold: less exposure to instructor expertise and learning through techniques to which these students are not accustomed.

It should be kept in mind that these results were limited to a quantitative business class, introductory business statistics. This study illustrates the need to measure empirically student performance outcomes and consider important student factors, namely grade point average, when performing research in the active learning domain.

APPENDIX
Bloom Taxonomy and Sample Assessment Questions

1. Binomial Distribution Topic

Knowledge/Comprehension Level

For a binomial distribution with the following characteristics:

- Probability of success 0.40
- 15 trials

Determine the probability of at least 5 successes.

Application Level

The increase or decrease in the price of a stock between the beginning and the end of a trading day is assumed to be an equally likely random event. Determine the probability that a stock will show an increase in its closing price in four out of five days.

Analysis/Evaluation Level

Skewness refers to lack of symmetry in a distribution. If a distribution is perfectly symmetrical, it is not skewed. If it is more likely that observations (or successes) will occur for the larger values in the distribution, the distribution is said to be left-skewed (or it is not very likely to obtain success for the smaller/left values in the distribution as compared to the larger/right values). The following distribution would be considered skewed in which way?

- A binomial distribution with probability of success 0.34 and 16 trials.

2. P-Values in Hypothesis Testing Topic

Comprehension Level

In a right-tail hypothesis test with test statistic $t^*=2.3$, $\alpha=.05$, and $n=11$, what is the appropriate statistical conclusion?

Application/Analysis Level

ATMs must be stocked with enough cash to satisfy customers making withdrawals over an entire weekend. If too much cash is unnecessarily kept in the ATMs, the bank is forgoing the opportunity of investing the money and earning interest. Suppose that in a random sample of 20 withdrawals over various weekends at a particular branch, the mean amount withdrawn was \$160 with a sample standard deviation of \$30. Assume withdrawal amounts are normally distributed. At the .05 level of significance ($\alpha=.05$) is there significant evidence to suggest that the mean withdrawal amount is more than \$150? If the computed test statistic came out to be 1.71 for a random sample of 25 withdrawals, how confident could we be in concluding the mean withdrawal was more than $150?

3. Sampling Distributions Topic

Application/Analysis Level

The diameters of ping-pong balls manufactured at a large factory are normally distributed. The population mean and standard deviation for the diameters are 1.30 and .055 inches, respectively. What is the probability the mean diameter of a sample of 30 balls will be between 1.29 and 1.31 inches? If the diameters were not normally distributed, what is the probability the mean diameter of a sample of 20 balls will be between 1.31 and 1.33 inches?

Analysis Level

If you were to select all possible samples of size 3 and set up the sampling distribution of the mean, how would this standard error compare to the standard error computed for samples of size 2? In other words, when we take larger sample sizes, what happens to the standard deviation in the sampling distribution?

References

Alsop, R. (2006). "Schools Find Fun A Worthy Teacher To Foster Creativity," Wall Street Journal - Eastern Edition, 248(61), B8.

Auster, E. R., & Wylie, K. K. (2006). "Creating Active Learning in the Classroom: A Systematic Approach," Journal of Management Education, 30(2), 333-353.

Barak, M., Lipson, A., & Lerman, S. (2006). Wireless Laptops as Means For Promoting Active Learning In Large Lecture Halls. Journal of Research on Technology in Education, 38(3), 245-263.

Becker, W. E. (1997). Teaching Economics to Undergraduates. Journal of Economic Literature, 35, 1347-1373.

Bloom, B. S. (1956). Taxonomy of educational objectives: The classification of educational goals. White Plains, NY: Longman.

Bonwell, C. C. (1997). Using active learning as assessment in the postsecondary classroom. Clearing House, 71(2), 73.

Bonwell, C. C., & Eison, J. A. (1991). Active learning : creating excitement in the classroom Washington, DC School of Education and Human Development, George Washington University.

Brown, R., & Guilding, C. (1993). A survey of teaching methods employed in university business school accounting courses. Accounting Education, 2(3), 211-218.

Dean, A. M., & Voss, D. (1999). Design and Analysis of Experiments: Springer.

Fai, A. H., & Cornelius, P. L. (1996). Approximate F-tests of multiple degree of freedom hypothesis in generalized least squares analysis of unbalanced split-plot experiments. Journal of Statistical Computing and Simulation, 54, 363-378.

Lee, K. H. (2007), “Chocolate Chip Cookies as a Teaching Aid,” The American Statistician, 61, 351-355.

Hansen, J. D. (2006). Using Problem-Based Learning in Accounting. Journal of Education for Business, 81(4), 221-224.

O'Brien, E. M., & Hart, S. (1999). Action Learning: The link between academia and industry? Educational Research, 41, 77-89.

Page, D., & Mukherjee, A. (2000). Improving undergraduate student involvement in management science and business writing courses using the seven principles in action. Education, 120, 547-557.

Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. Biometrika, 58(545).

Raelin, J., & Coghlan, D. (2006). Developing Managers as Learners and Researchers: Using Action Learning and Action Research. Journal of Management Education, 30(5), 670-689.

Revans, R. (1983). Action Learning, Its Terms and Character. Management Decision, 21(1).

Revans, R.(1971). Developing effective managers; a new approach to business education: Praeger Publishers.

Sarason, Y., & Banbury, C. (2004). Active Learning Facilitated by Using a Game-Show Format Or Who Doesn't Want to be a Millionaire? Journal of Management Education, 28(4), 509-518.

Schneider, M. (2001, March 26). MBAs for working stiffs: Growing demand has schools beefing up part-time programs. Business Week.

Sutherland, T. E., & Bonwell, C. C. (1996). Using Active Learning in College Classes: A Range of Options for Faculty (Vol. 67).

Ueltschy, L. C. (2001). An Exploratory Study of Integrating Interactive Technology into the Marketing Curriculum. Journal of Marketing Education, 23(1), 63-72.

Umble, M., & Umble, E. J. (2004). Using Active Learning to Transform the Monte Hall Problem into an Invaluable Classroom Exercise. Decision Sciences Journal of Innovative Education, 2(2), 213-217.

Verbeke, G., & Molenberghs, G. (1997). Linear Mixed Models in Practice. New York: Springer-Verlag.

West, B., Welch, K., & Galecki, A. (2007). Linear Mixed Models: A Practical Guide Using Statistical Software. Boca Raton: Chapman & Hall.

David Weltman
Department of Information Systems and Operations Management
The University of Texas at Arlington
Arlington, TX 76019
E-mail: *dweltman@uta.edu*

Mary Whiteside
Department of Information Systems and Operations Management
The University of Texas at Arlington
Arlington, TX 76019
E-mail: *whiteside@uta.edu*
