Effect of Class Size on Student Outcomes in Mathematics Courses with Technology Assisted Instruction and Assessment

Jim Gleason
The University of Alabama

Abstract: The implementation of online texts, videos, homework, and tests has changed the process of instruction in introductory college mathematics courses. With this change, more of the students' learning takes place outside of the traditional college classroom and in places such as tutoring centers and dorm rooms. This study explores how these changes influence the impact of the size of the classroom portion of the learning experience on student learning, student achievement, student engagement, and student satisfaction. A combination of chi-square tests for independence with unordered categorical data and Mann-Whitney two-sample rank-sum tests for continuous data and ordered categorical data are used to analyze student outcomes generated from College Algebra and Applied Calculus courses with class sizes ranging from 37 to 129 with common syllabi, homework, quizzes, and tests. These tests showed no significant difference in student learning or student achievement and had mixed results regarding student engagement and student satisfaction.

Keywords: introductory mathematics, classroom-teaching experiment, class size

Introduction

With reduction in state funding at most public universities, there is increased pressure upon the universities to reduce costs. One of the initial locations to look for these cost reductions is an enlargement of class sizes, thereby increasing the number of student-hours for which each faculty member is responsible. While this may make sense financially, it may have a negative effect upon student retention and student graduation, two areas of continued interest among university administration.

In order to counteract the negative impact of large class sizes, many schools look to technology to provide the solutions. Personal response systems (or clickers) have proven to be very popular at improving student engagement in the classroom (Judson & Sawada, 2002; Caldwell, 2007), but still need more research to determine the true impact (Fies & Marshall, 2006). Another direction is to use instructional software packages to reduce the importance upon the lecture part of the course. Early results are positive with improved pass rates, college retention, and student satisfaction (Twig, 2003).

Background

Over the years, many researchers have studied the impact on large class sizes at universities on many different factors involved in student achievement and satisfaction. These include student perceptions, level and amount of learning, and the level of interaction between the instructor and the student.

One current area of emphasis among colleges and universities is improving student retention rates, the percentage of students continuing from their first to second years at the institution.

Since a small number of courses generate a large number of student credit hours, with most of these courses included in the first year of a program of study and in large class sizes, the outcome from these few courses can have a significant impact on student retention (Twig, 2003). While the size of a single class likely has little direct effect on student retention rates, there is likely an indirect effect through student achievement and student satisfaction.

The relationship between class size and student achievement is complex with many conflicting results. While some studies show that class-size has negligible impact on student achievement (Williams, Cook, Quinn, & Jensen, 1985; Pascarella & Terenzini, 1991), the majority demonstrate an inverse relationship in that as the class size increases, student achievement declines (Kokkelenberg, Dillon, & Christy, 2008; Franklin & Theall, 1991; Light, 2001; Lindsay & Paton-Saltzberg, 1987). Even among the studies demonstrating this inverse relationship, many support the theoretical model put forth by Glass, Smith and their colleagues (Glass & Smith, 1979; Glass, Cahen, Smith, & Filby, 1982), through meta-analysis of the research, that the function mapping class-size to student achievement is decreasing and concave up (Cuseo, 2007; Kokkelenberg, Dillon, & Christy, 2008). In other words, the impact of class size upon student achievement decreases as the class size increases.

In addition to the effect on student achievement, research supports an effect of large classes upon student satisfaction. Using a fixed-effects model, Bedard and Kuhn (2008) conclude that there is a negative non-linear relationship between class size and student evaluations stronger than the relationship to student achievement, and with less concavity. This supports earlier findings including an analysis of studies which revealed a similar negative relationship between class size and student evaluation, particularly in regards to instructor interactions with students (Feldman, 1984).

In an analysis of the research, Cuseo (2007) divided into seven categories the areas in which there is a negative impact of larger class sizes upon the student learning experience. These seven categories include an increased faculty reliance on the lecture method of instruction, a reduction of students' active involvement in the learning process, a reduction of frequency and quality of instructor-student interactions, less depth of thinking inside the classroom, a limit of breadth and depth in objectives and assignments, lower academic performance, and less course satisfaction. In the current study, we control for all but the academic performance, students' involvement in the learning process, and course satisfaction through a very rigid structure within each of the two courses by having common assignments, quizzes, tests, requirements for instructor feedback, common lecture material, and an overlap in instructors in both the control and experimental classrooms.

Methodology

Courses Studied

This study focuses on how the use of instructional software packages, computer labs with tutoring, and increased electronic student-teacher interaction influences the effects of large class sizes upon student achievement and engagement. Each of the two classes studied, College Algebra and Applied Calculus, are three semester unit courses. Applied Calculus met three academic hours a week in a classroom in addition to the computer assisted portion of the course and College Algebra met one academic hour a week with two required hours in the computer and tutoring center.

This computer-assisted portion involved all sections of the class having common syllabi, schedule, homework assignments, quizzes, and tests. All of these assessment components took

place using computer software that included student assistance modules on homework assignments such as videos, e-textbook, and examples. The university also provided a 400-seat computer and tutoring center staffed with instructors, graduate students, and undergraduate tutors 71 hours each week to help students with their mathematics learning. It was in this computer lab that all tests took place, with homework and quizzes accessible over the internet.
For the analysis, classes are categorized as medium (between 30 and 55), and large (between 110 and 130), with approximately equal distribution between classes that met for 75 minutes each on Tuesday and Thursday and those that met for 50 minutes on Monday, Wednesday, and Friday (See Table 1).

Table 1: Distribution of Students in Courses

	Medium Size (30-55 students)	Large Size (110-130 students)	Total
College Algebra			
Tuesday/Thursday	547	231	778
Monday/Wednesday/Friday	402	447	849
Total	949	678	1627
Applied Calculus			
Tuesday/Thursday	180	130	310
Monday/Wednesday/Friday	172	128	300
Total	352	258	610

Dependent Variables

To measure student learning, we used the students' final exam scores. For each of the courses, the final exam consisted of 50 multiple-choice questions on material from the entire semester. These exams were common across sections with the students taking their exams during times of their choosing throughout the final exam week in the large computer lab. Since almost all students completed the test in the time given, and since there was no penalty for guessing, we used a 3-parameter item response theory model with MULTILOG (Thissen, 2003). The College Algebra final exam had a marginal reliability of 0.8726 ($n = 1438$, negative twice the log likelihood of 40961.9) and the Applied Calculus final exam had a marginal reliability of 0.8828 ($n = 524$, negative twice the loglikelihood of 17597.9), which demonstrated that the scores on the final exam had a strong discrimination at the group level and were appropriate for our analysis. Furthermore, as can be seen from Figure 1, this distinguishing ability was strongest for the students who pass the final exam.

Since the final exam scores were skewed left, they did not satisfy the normality condition necessary to use a parametric test. Thus, for the analysis of the final exam scores we used both the Mann-Whitney two-sample rank-sum test, since the scores dispersed enough that individuals could be ranked with a relatively limited number of ties, and the Kolmogorov-Smirnov test. The Mann-Whitney two-sample rank-sum test is similar to Student's t-test, but without the normality assumption and has an asymptotic efficiency of ${}^{3}/_{\pi} \approx 0.95$ when normality holds (Higgins, 2004, p. 63). The Mann-Whitney test assumes that the two distributions have a similar variance, while the Kolmogorov-Smirnov test has no assumption on the distributions and compares the cumulative distribution functions for the two populations.

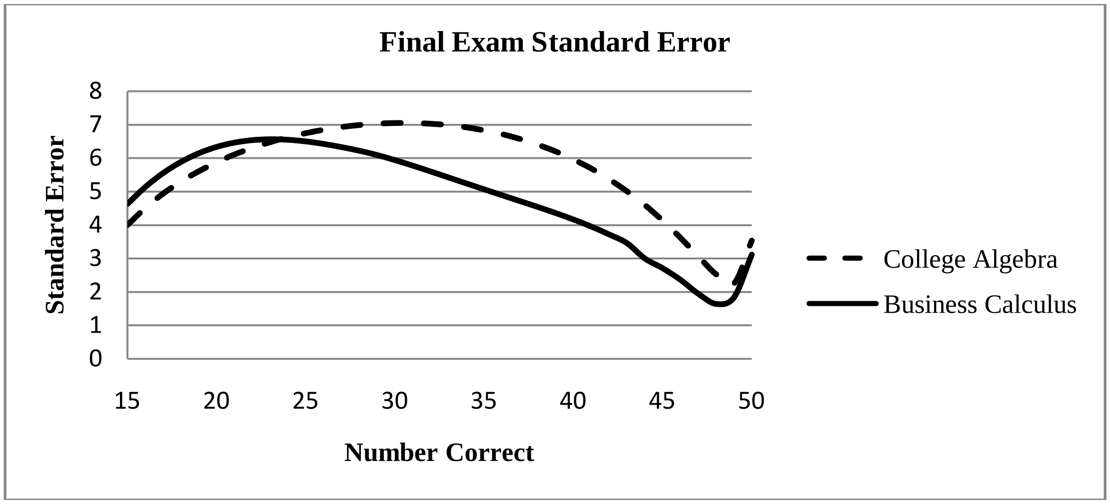

Figure 1: Standard Error for Final Exams

In addition to using the students' final exam scores, the analysis also included the students' course points to measure achievement in the course. While there was a strong correlation between students' final exam score and their course points, the final exam score focused on the students' subject knowledge while the course points included factors, such as work ethic measured through attendance, homework completions, quizzes, etc. Additionally, the analysis for the course points included all students enrolled for the course, while the procedure for the final exam grade included only those students who took the final exam. Since the student course points were bimodal and dispersed, a Mann-Whitney two-sample rank-sum test and a Kolmogorov-Smirnov test were appropriate for the analysis.

For the College Algebra course, the university had a policy where the students could receive grades of A, B, or C (with plus and minus options). If the student did not receive one of these passing grades, they either have withdrawn from the course and received a W or received no credit for the course and a NC on their transcript that does not count toward their grade point average. For the Applied Calculus course, the NC option was no longer available and students received a grade of A, B, C, D, or F (with plus and minus options available on all but the F) or the student had withdrawn from the course.

In the analysis, we converted the grades into the standard 4-point scale, adjusting for plus and minus grades by adding or subtracting 0.33 respectively, and with W, NC, and F all counting as a 0. An additional analysis studied the pass rate by marking those students with any type of A, B, or C with a 1 and the others with a 0. For each of these analyses, we used a chi-square test for independence to determine if the distribution of grades and passing had the same shape for both the large classes and the medium classes. For the analysis of the scaled grades, we also used a Mann-Whitney two-sample rank-sum test to analyze the singly ordered contingency table since this test would better detect the influence of class size (Higgins, 2004, pp. 176-178).

Another desired student outcome is student engagement in the course through active learning. Since homework was assigned and completed on the computers almost every class period, the homework completion percentage and the homework grade for the course related strongly to how much active learning with which the student is engaged. With a large amount of emails from faculty members reminding students to complete their homework and with learning aides available while students work on the homework, a large number of students attempted all of the homework assignments (35% for College Algebra and 33% for Applied Calculus) and received perfect homework scores (14% for College Algebra and 11% for Applied Calculus).

Since there is little difference between students who attempted 95% of their homework and those who attempted 100% of the homework in terms of student engagement we did not wish to distinguish between these levels. We could not use any parametric statistics due to the large difference from normality of the distribution and we cannot use a rank-sum test using the actual attempt percentages as that would have created a major difference between those who attempted 95% of their assignments and those who attempted all of their assignments. Therefore, we classified the homework attempt percentage and scores into four categories which we labeled as disengaged (less than 70%), slightly engaged (greater than or equal to 70% but less than 80%), moderately engaged (greater than or equal to 80% but less than 95%), and fully engaged (greater than or equal to 95%). While these categories are somewhat arbitrary, they arose from the distribution of values, taking into account the relationship with the impact upon students' grades, i.e. if a student completes less than 70% of the homework, the homework grade was less than a passing grade and took away from the pass rate. In each class, for both homework attempts and scores, the most students fell into the fully engage category followed by the moderately engaged and disengaged, with the slightly engaged category having very few students (less than 10%). Since the contingency table in singly ordered, we used a Mann-Whitney two-sample rank-sum test to analyze the effect of the class size.

Another method used to measure student engagement involved student responses to an end of course survey. On the survey, students were asked two related questions with a 5-point Likert scale. The questions "Did you come to class prepared by having completed the assignments?" and "How frequently did you attend class?", had options of always, usually, often, sometimes, and rarely. These two questions built upon the other data measuring student engagement by including the students' perspectives. We analyzed this data using a Mann-Whitney two-sample rank-sum test.

To evaluate the impact of class size upon student satisfaction, we used a Mann-Whitney two-sample rank-sum test to analyze student responses on three questions from the college-wide student evaluation of instruction. These three questions used a 5-point Likert scale and included, "The course was a valuable learning experience" (strongly agree, agree, undecided, disagree, strongly disagree), "How much did you learn in this course?" (a great deal, much, some, little nothing), and "How would you rate this course?" (excellent, above average, average, below average, failure).

Results

Student Knowledge

Using the students' final exam scores as a measure of their knowledge at the end of the course, we compared the students in the large classes with the students in the medium classes using Mann-Whitney two-sample rank-sum and Kolmogorov-Smirnov tests. Using the Mann-Whitney test, for both the College Algebra ($U = 247808.5, p = 0.620$) and Applied Calculus ($U = 30927.5, p = 0.148$) courses, there was no significant difference between the two distributions and so we could not reject the null hypothesis. For the College Algebra course, the mean rank for the medium classes ($n = 856$) was 727, and 716 for the large classes ($n = 588$). For the Applied Calculus course, the mean rank and for the medium classes ($n = 305$) was 254, and 274 for the large classes ($n = 219$). The Kolmogorov-Smirnov tests also show no significant differences between the two distributions for either College Algebra (K-S $= 0.593, p = 0.873$) or Applied Calculus (K-S $= 1.011, p = 0.258$).

Student Achievement

Using the students' points from each of the courses, the Mann-Whitney two-sample rank-sum tests showed no significant difference between the medium class and large classes in either College Algebra ($U = 304628, p = 0.67$) or Applied Calculus ($U = 43897.5, p = 0.482$) and so we could not reject the null hypothesis. In the College Algebra course, the mean rank for students in the medium classes ($n = 949$) was 832, and in the large classes ($n = 678$) was 789. For the students in the Applied Calculus course, the mean rank in the medium classes ($n = 352$) was 301, and was 311 in the large classes ($n = 258$). Similarly, the Kolmogorov-Smirnov tests showed no significant difference for College Algebra (K-S $= 1.173, p = 0.128$) or Applied Calculus (K-S $= 0.650, p = 0.792$).

Student Grades and Pass Rates

When comparing student grades based on a four-point scale with plus and minus, there was no significant difference, using the chi squared test, between the large and small classes in both College Algebra ($\chi^2(9, n = 1627) = 4.85$, $p = 0.8468$) and Applied Calculus ($\chi^2(9, n = 610) = 8.05$, $p = 0.7084$). There was also no difference when using the Mann-Whitney two-sample rank-sum test for either College Algebra (medium class mean rank = 829.45, large class mean rank = 792.37,$U = 307046, p = 0.111, n = 1627$) or Applied Calculus (medium class mean rank = 300.89, large class mean rank = 311.78,$U = 43787, p = 0.447, n = 610$), although as expected, the Mann-Whitney test was more powerful at finding differences. Therefore, we were not able to reject the null hypothesis.

Similarly, when comparing pass rates for the courses we were unable to reject the null hypothesis that there is no significant difference between the two class sizes for either College Algebra ($\chi^2(1, n = 1627) = 3.255$, $p = 0.0712$) or for Applied Calculus ($\chi^2(1, n = 610) = 0.5290$, $p = 0.4670$). In particular, the pass rate was higher for medium size classes in the College Algebra course (74% versus 70%) and lower for the medium size classes in Applied Calculus (68% versus 71%).

Student Engagement

We measured student engagement using the two variables of average homework score and homework attempt percentage. We then categorized these students into four categories for each variable; disengaged, slightly engaged, moderately engaged, and fully engaged. The number in each category of the homework scores are given in Table 2 while the distributions for the homework attempts are in Table 3.

In the College Algebra course, the Mann-Whitney two-sample rank-sum test measured a significant difference between the medium and large sized classes in both homework attempts and homework grades. The analysis, using the categorization from homework attempts, revealed that the medium sized class ($n = 949$) had a mean rank of 843.97 with the large class ($n = 678$) having a mean rank of 772.06 to give a Mann-Whitney U of 293273 ($p = 0.001$). The categorization from homework scores gave the medium sized class a mean rank of 833.84, the large class a mean rank of 786.24, and a Mann-Whitney U of 302887 ($p = 0.031$).

Table 2: Homework Score Distributions

	College Algebra		Applied Calculus	
	Large	Medium	Large	Medium
Disengaged	130	148	50	68
Slightly Engaged	51	66	20	22
Moderately Engaged	213	295	84	93
Fully Engaged	284	440	104	169
Total	678	949	258	352

Table 3: Homework Attempt Distributions

	College Algebra		Applied Calculus	
	Large	Medium	Large	Medium
Disengaged	120	126	50	64
Slightly Engaged	65	78	13	23
Moderately Engaged	173	224	76	76
Fully Engaged	320	521	119	189
Total	678	949	258	352

The analysis of the Applied Calculus course on the other hand did not show a significant difference between the two sizes. Even though the medium sized classes ($n = 352$) outperformed the large classes ($n = 258$) in both the homework attempt mean rank (313.12 to 295.10) and homework score mean rank (313.57 to 294.48), the homework attempt analysis gave a Mann-Whitney U of 42724 ($p = 0.176$) and the homework score analysis gave a Mann-Whitney U of 42566 ($p = 0.158$). So we could not reject the null hypothesis for this course.

In addition to the analysis using homework attempts and scores, the student engagement was also measured from the student perspective using end of course surveys. Using two items with five-point Likert scales, neither the College Algebra course nor the Applied Calculus course demonstrated a significant difference in student perception of their engagement in either the medium or large classes.

For the question regarding coming to class having completed the assignments, the medium College Algebra classes ($n = 508$) had a mean rank of 411.34 while the large classes ($n = 307$) had a mean rank of 402.47 ($U = 76279, p = 0.555$). Similarly, in the Applied Calculus course, the medium classes ($n = 150$) had a mean rank of 132.9 and the large classes ($n = 105$) had a mean rank of 121 ($U = 7139.5, p = 0.137$).

For the question regarding frequency of attendance the medium sized ($n = 509$) College Algebra classes had a mean rank of 406.78 and the large classes ($n = 307$) had a mean rank of 411.34 ($U = 77258.5, p = 0.760$). The same question in the Applied Calculus course recorded the medium classes ($n = 150$) with a mean rank of 127.99 and the large classes ($n = 105$) a mean rank of 128.01 ($U = 7874, p = 0.998$).

Student Satisfaction

To measure student satisfaction, we analyzed the responses to three five-point Likert scale items on the end of course evaluations. For the College Algebra course, the large classes outperformed the small classes in all three items. For the item, “The course was a valuable learning experience," the large class ($n = 306$) had a mean rank of 423.64 compared to a mean rank of 395.31 for the medium class ($n = 505$) ($U = 71868.5, p = 0.077$). The item, "How much did you learn in this course?", the mean rank for the large class ($n = 307$) outperformed the medium class ($n = 509$) 432.06 to 394.29 ($U = 70898, p = 0.020$). When the students were asked to rate the course, the large classes ($n = 305$) again surpassed the medium classes ($n = 509$) by 418.22 to 401.08 ($U = 74354, p = 0.292$). While the results of only one of the three items reached statistical significance, the overall result is that students seem to prefer the larger classes.

The results for the Applied Calculus course matched those of the College Algebra course with the mean rank of large classes ($n = 105$) outranking the medium classes ($n = 150$) in the value of the learning experience (133.18 to 122.69, $U = 7105.5, p = 0.224$), the amount learned in the course (142.55 to 117.82, $U = 6347.5, p = 0.005$), and in the rating of the course (139.41 to 120.01, $U = 6676.5, p = 0.030$). Thus, we conclude that the students who responded to the survey in the large classes had a higher view of the course than those who responded in the medium classes with two of the items achieving statistical significance.

Discussion

As with much of the literature, the results in this study were mixed, with most of the variables analyzed not showing any statistical significance. This lack of difference between the medium sized classes and the large classes likely relates to one of two main factors. Either the extensive use of technology in the course to encourage students to remain active learners through frequent homework, quizzes, and tests overcame the difficulties of the larger class sizes or that the difference between the class sizes is negligible since the greatest measured effects of class size occur with classes less than 30 students (Kokkelenberg, Dillon, & Christy, 2008).

The area that showed the most negative effect of increased class size was student engagement as measured by homework attempts and homework scores. This fit within the assumption that students in larger classes tend to feel more anonymous and are therefore more likely to not stay engaged with the material. What is interesting is that this lack of engagement seemed to have limited effect on student achievement, seemingly contradicting the commonly held belief that in order for students to succeed in a course they must be constantly engaged in the course.

Further analysis of the data on student engagement revealed that the difference in the distribution of students in the medium and large classes of College Algebra did not occur in the middle two categories, but between the disengaged and the fully engaged. This difference in distributions pointed toward a shift of 5% the population in the large classes shifting down in each category leaving an extra 5% in the bottom category. This implied that the lack of engagement is not localized among one particular type of student but affects students at all engagement levels. Therefore, when dealing with a large class, the instructor must keep this in mind and pay particular attention to keeping students on task, giving the students a sense of ownership in the course, and reminding them to work on the homework.

The most probable reason that this effect was not as strong in the Applied Calculus course is that with the students meeting with the instructor for three academic hours per week instead of one, they likely felt more ownership in the course. Another possible explanation is that the students in Applied Calculus had already completed College Algebra and were therefore more likely to have a higher maturity level and act more responsible in regards to the homework assignments.

The most surprising result in the study involved the student responses to the end of course evaluations. That more students in the large classes felt that the course was a valuable experience, they learned more, and thought highly of the course seemed to contradict previous studies showing that larger classes had a negative impact on student satisfaction and retention (Cuseo, 2007). One possible factor involved is that students in the medium size classes responded to the student survey at a slightly higher rate than the students in the large classes did (59% to 52% in College Algebra and 49% to 48% in Applied Calculus). Since the responses most likely coming from the more engaged students, a larger percentage of disengaged students likely responded to the survey from the medium sized classes. Another hypothesis relates to a potential difference in expectations of students dependent upon the size of the classroom. For example, students in a smaller class would have a higher expectation of learning more material, that the instructor would do a better job of imparting the course content, or that the instructor would care more about the students. Therefore, if the class size does not affect these areas, the student evaluations for the larger classes would be higher than those of the smaller classes.

In conclusion, while there was some measurable difference between the two sizes, the difference did not always favor the smaller classes and the impact was minimal upon the factors that have the greatest impact on long term student success, namely student grades and student satisfaction. Therefore, the results of this study imply that changing classes from 30-50 students to classes with 110-130 students does not negatively impact the student experience when connected with a heavy component involving computer assessment and frequent instructor interaction with the students via email.

Bibliography

Bedard, K., & Kuhn, P. (2008). Where class size really matters: Class size and student ratings of instructor effectiveness. *Economics of Education Review , 27,* 253-265.

Caldwell, J. (2007). Clickers in the large classroom: Current research and best-practice tips. *CBE Life Sciences Education , 6* (1), 9-20.

Corder, G. W., & Foreman, D. I. (2009). *Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach.* Hoboken, NJ: John Wiley & Sons, Inc.

Cuseo, J. (2007). The empirical case against large class size: Adverse effects on the teaching, learning, and retention of first-year students. *The Journal of Faculty Development , 22* (1), 5-21.

Feldman, K. A. (1984). Class size and college students' evaluations of teachers and courses: A closer look. *Research in Higher Education , 21* (1), 45-116.

Fies, C., & Marshall, J. (2006). Classroom response systems: A review of the literature. *Journal of Science Education and Technology , 15* (1), 101-109.

Franklin, J. L., & Theall, M. (1991). Grade inflation and student ratings: A closer look. *Paper presented at the 72nd Annual Meeting of the American Educational Research Association.* Chicago, IL.

Glass, G. V., & Smith, M. L. (1979). Meta-analysis of research on the relationship of class-size and achievement. *Evaluation and Policy Analysis* (1), 2-16.

Glass, G. V., Cahen, L. S., Smith, M. L., & Filby, N. N. (1982). *School class size, research and policy.* Beverly Hills: Sage Publications.

Higgins, J. J. (2004). *Introduction to Modern Nonparametric Statistics.* Pacific Grove, CA: Brooks/Cole.

Judson, E., & Sawada, D. (2002). Learning from past and present: electronic response systems in college lecture halls. *The Journal of Computers in Mathematics and Science Teaching*, *21* (2), 167-181.

Kokkelenberg, E. C., Dillon, M., & Christy, S. M. (2008). The effects of class size on student grades at a public university. *Economics of Education Review* , *27*, 221-233.

Light, R. J. (2001). *Making the most of college: Students speak their minds.* Cambridge, MA: Harvard University Press.

Lindsay, R., & Paton-Saltzberg, R. (1987). Resource changes and academic performance at an English Polytechnic. *Studies in Higher Education* , *12* (2), 213-227.

Lunney, G. H. (1970). Using analysis of variance with dichotomous dependent variable: An empirical study. *Journal of Educational Measurement* , *7* (4), 263-269.

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students.* San Francisco: Jossey-Bass.

Thissen, D. (2003). MULTILOG for Windows (version 7.0) [Computer Software]. Mooresville, IN: Scientific Software International, Inc.

Twig, C. A. (2003). Improving quality and reducing cost: Designs for effective learning. *Change* , *35* (4), 23-29.

Williams, D. D., Cook, P. F., Quinn, B., & Jensen, R. P. (1985). University class size, is smaller better? *Research in Higher Education* , *23* (3), 307-317.