

Using a Quasi-Experimental Design in Combination with Multivariate Analysis to Assess Student Learning

Michael Delucchi
University of Hawaii
delucchi@hawaii.edu

Abstract: College professors have adopted numerous strategies for teaching undergraduates, yet few researchers provide empirical evidence students' learning actually increased because of the instructional innovation. Assessment of pedagogy is frequently subjective and based on comments from students or faculty. Consequently, evaluating the effectiveness of teaching activities on college student learning, in general, and in statistical analysis courses, in particular, is warranted. This study employed a pretest-posttest design to measure student learning and then examined the relationship between student demographics, prior knowledge, and course characteristics on knowledge gained in undergraduate statistics. Data derived from 185 students enrolled in six different sections of a statistical analysis course taught over a seven-year period by the same instructor. Multiple regression analyses revealed age, age X gender (interaction effect), major, prior knowledge, examinations, and group projects all had statistically significant effects on how much students learned in the course. The results suggest faculty assess students' prior knowledge at the beginning of the semester and use such data to inform both the content and delivery of statistical analysis. Moreover, before embracing a new pedagogy, faculty should establish empirically that learning is linked to the teaching innovation.

Keywords: learning versus performance, knowledge gain, direct assessment, pretest/posttest design, multivariate analysis, groupwork, teaching statistic.

Coursework in statistical analysis is required in most undergraduate social science and professional studies programs. Despite the importance of quantitative skills in these disciplines, students with insufficient preparation in algebra and/or math anxiety confront faculty teaching statistics (Bandalos, Finney, & Geske, 2003; Blalock, 1987; Cerrito, 1999; Forte, 1995; Garfield and Chance, 2000; Macheski et al., 2008; Wilder, 2010). In response, professors assigned to statistical analysis courses search for pedagogy to reduce student anxiety and increase quantitative knowledge. Many instructional techniques (e.g., small-group work, collaborative testing, humor, computer-assisted data analysis, active learning, etc.) have been recommended for teaching undergraduates (Delucchi, 2006; Helmericks, 1993; Schacht and Stewart, 1990; Schumm et al., 2002; Strangfeld, 2013). Faculty employing these practices report greater student satisfaction with the course (Fischer, 1996; Perkins and Saris, 2001; Potter, 1995; Stork, 2003), reduced math anxiety (DeCesare, 2007; Lomax and Moosavi, 2002), and a belief student learning was greater than could have been achieved without the teaching innovation (Auster, 2000; Wybraniec and Wilmoth, 1999; Yamarik, 2007).

While not without some value, most studies offer little direct empirical evidence that student's knowledge, i.e., learning, increased as a result of pedagogy. Assessment of learning tends to rely on student comments or faculty impressions (Fisher-Giorlando, 1992; Lomax and Moosavi, 2002; Marson, 2007; Schacht and Stewart, 1992). Perceptions of learning and even quantitative student evaluations of teaching (SETs) do not represent direct measurement of learning. As indicators of *perceived* knowledge (rather than actual knowledge), these indirect assessments of learning are limited by assumptions that must be made about what such self-reports constitute (Price and Randall, 2008).

Often used as proxies for learning, students' quiz, examination, and course grades (Borresen, 1990; Delucchi, 2007; Perkins and Saris, 2001; Smith, 2003; Yamarik 2007) do not represent direct

indicators of learning (Baker, 1985; Chin, 2002; Garfield and Chance, 2000; Lucal et al., 2003; Luce & Kirnan, 2016; Wagenaar, 2002; Weiss, 2002). Grades measure academic performance. Learning is increased knowledge, i.e., the difference between what students know at the beginning of the course, compared to the end of the semester. Performance is demonstrating mastery, e.g., accurate statistical computation or correct responses to multiple-choice items on examinations. Cognitive psychologists have found that taking a quiz or examination improves students' subsequent learning of course material. How this occurs is not clear, but testing appears to make ensuing study more effective (Little and Bjork, 2010; Little and Bjork, 2011). Therefore, examinations may function as teaching devices that motivate students to study course content (Bjork and Bjork, 2011). This is noteworthy, as researchers consistently report a positive association between learning and studying (Arum and Roksa, 2011; Astin, 1993), i.e., the more time students spend studying, the more they learn.

The distinction between learning and performance is important, because students enter courses with unequal knowledge, skills, and academic experiences. For example, an individual may begin the semester knowing little, learn a great deal, perform adequately, and receive an average grade, or a student may enter a course knowing a great deal, learn a small amount, perform very well, and earn a high grade (Neuman, 1989). While such students' examination scores and course grades represent performance, neither is a direct assessment of learning. Rather, a pretest is required at the beginning of a course (to establish students' prior knowledge) and a posttest is necessary to measure learning (i.e., knowledge gain) after the course is completed (Luce & Kirnan, 2016).

Quasi-experimental research on student learning in university-level statistics courses is rare (Bridges et al., 1998; Luce & Kirnan, 2016; Price & Randall, 2008). For that reason, the purpose of this study is twofold. First, to assess learning, a pretest-posttest design is used to compare students' statistical knowledge at the beginning of the course with measurement at the end of the semester. Second, multivariate analysis is employed to examine the effects of individual attributes (e.g., age, gender, major, prior knowledge) and course characteristics (e.g., time of day, class size, group projects, quizzes, and examinations) on student learning.

Data and Methods

The study was conducted at a small (approximately 2,500 students), state-supported baccalaureate degree granting university in the United States. The "Carnegie Classification for Institutions of Higher Education" describes the university as a Baccalaureate College: Diverse Fields (Center for Postsecondary Research, 2015). The institution is co-educational (66% women; 34% men), ethnically diverse (59% ethnic minorities), and comprised of many nontraditional age (30% 25 years of age or older) students. Eighty-two percent of the student population is employed (40% working more than 31 hours per week), and all students commute to the campus.

Course Description

Statistical Analysis is an undergraduate course taught in the Division of Social Sciences that serves as an introduction to descriptive and inferential statistics. Completion of algebra II (or a higher-level mathematics course) with a grade of "C" or better is the prerequisite. *Statistical Analysis* is required for all social science majors (e.g., anthropology, economics, political science, psychology, and sociology) at the university. In addition, the course can be taken to fulfill a core requirement for some professional studies majors (e.g., early childhood education, health care administration, justice administration, and public administration). As a result, approximately 70 percent of the students enrolled in *Statistical Analysis* are social science majors and 30 percent come from other programs. Course requirements included three examinations, i.e., Examination 1 (15%), Examination 2 (20%),

Final Examination (35%), two small group projects worth 10% each, and twelve quizzes weighted a combined 10%. Computational problems and computer exercises using the Statistical Package for the Social Sciences (SPSS) were assigned from the textbook, but not graded.

Sample

Student data derived from class records for six sections of *Statistical Analysis* taught over a seven-year period. Complete information was obtained for 185 of the 214 students enrolled in the course at the beginning of each semester, representing an 86% response rate. The class met for 80 minutes, twice a week, during a fifteen-week semester. Course content, delivered via lectures and class discussions, paralleled chapters in the text. While the text, most recently Healey (2015), changed as new editions became available, the instructor, lectures, homework, quizzes, group projects, examinations, and grading criterion were essentially constant across the six sections of the course.

Measures

Pretest-Posttest Instrument. To assess students' statistical knowledge, a comprehensive multiple-choice test was developed and administered at the second-class meeting during the first week of the semester.¹ This pretest contained 30 questions on descriptive and inferential statistics derived from "typical" computational and quantitative reasoning skills covered in the *Statistical Analysis* course. (See Appendix for pretest-posttest content areas.) The same instrument was administered as a posttest at the last scheduled class session. Students were given 50 minutes to complete each test and could use a calculator and consult their textbook. Pretest and posttest scores did not count toward students' course grade.

Only students who completed both tests were included in the data set. The Office of Institutional Research and Assessment (serving as the campus institutional review board for faculty research using student and course-level data) approved the *Statistical Analysis* course pretest-posttest project upon which this study is based. Information collected and analyzed did not include student names or any individual identifiable information.

Dependent Variable. In this study, the term "learning" refers to improvement over the 15-week semester in measurable statistical analysis skills and knowledge. The dependent variable (Improvement) measured learning or knowledge gained from the course. Improvement was calculated by subtracting the percentage of correct answers (out of 30) students received on the pretest from the percentage correct on the posttest. Positive values represented an increase in students' statistical knowledge from the beginning to the end of the course (Posttest percentage – Pretest percentage = Improvement, i.e., learning), while "0" or negative percentages represented no improvement. The higher the percentage, the more student knowledge gained or material learned.

Independent Variables. In addition to the pretest and posttest, students completed three examinations during the semester. These tests required students to perform statistical computations and to interpret their results. During the 80-minute class period, students worked independently, but were permitted use of a calculator, textbook, lecture notes, quizzes, homework, and group projects.

¹This test was created by sampling content from materials used in *Statistical Analysis*, including homework exercises, quizzes, examinations, projects, and textbooks. The instrument was "pilot" tested in a *Statistical Analysis* course one semester prior to its implementation in the study. Based on feedback from students and their performance, the test was revised, primarily to clarify specific questions.

The arithmetic average of the three examinations, each coded on a 0 to 100-point scale, served as an independent variable (i.e., Exam Mean).

Approximately once a week during the final 10-15 minutes of class, students were administered a quiz. Each quiz involved computations and interpretations similar to (but less rigorous than) those on examinations. Students could use a calculator, textbook, lecture notes, and their homework, but were required to complete quizzes independently. The first four quizzes covered descriptive statistics and corresponded to quantitative skills assessed on Examination 1. Quizzes 5 thru 8 focused on inferential statistics and represented content evaluated on Examination 2. The last four quizzes addressed statistical relationships and required knowledge similar to that on the Final Examination. The arithmetic average of the twelve quizzes, scored on a 0 to 10-point scale, was computed and used an independent variable (i.e., Quiz Mean).

Course requirements also included the completion of two group projects. Approximately four weeks prior to a projects' due date, students were instructed to organize themselves into two to four-member groups.² Groups decided how to divide the workload, but each member was required to be involved in all stages of the project. Students were collectively responsible for their project and all members received a group grade. To discourage "free riders" (i.e., individuals who contribute little or nothing the project), students were asked to apprise the professor if some members did not attend group meetings or were not performing their share of responsibilities. After the initial formation of the groups, students met outside of class. Groups were encouraged to meet with the instructor when they had questions and to submit rough drafts of their papers.

Group Project 1 introduced students to material that would appear on Examination 1. Working together, students used SPSS to compute frequency distributions, cross-tabulations, and descriptive statistics (i.e., measures of central tendency and dispersion) for nominal, ordinal, and ratio scale variables. After obtaining an SPSS printout, the group was required to interpret the data and write up the results in a two to three-page paper. Group Project 2 included content (e.g., correlation and regression) found on the Final Examination. Groups were required to select one scholarly article on reserve in the university library. Each group was instructed to discuss their article and interpret its findings. Subsequently, the group was required to compose a two to three-page paper demonstrating their ability to interpret multiple regression, as it appeared in the article. The arithmetic average of grades (assigned on a 0 to 12-point scale) awarded on Group Project 1 and Group Project 2 served as an independent variable (i.e., Group Projects).

Additional Independent Variables: Individual characteristics included student age, gender, major, and prior knowledge (percentage of correct answers on pretest). Class size and course meeting time were also recorded. Table 1 presents coding information and descriptive statistics for the dependent and all independent variables used in the study.

Table 1. Variables, Indicators, Means, and Standard Deviations (N = 185)

Variable	Indicator	Mean	S.D.
Pretest	Arithmetic average of the percentage of correct answers on the pretest.	43.89	11.13
Posttest	Arithmetic average of the percentage of correct answers on the posttest.	64.76	12.99
Improvement	Difference between the percentage correct on the pretest and posttest.	20.88	11.55

² Project groups ranged in size from two to four members, unfortunately, I did not collect data on the exact size of each group. Consequently, I was unable to control for the effects of group size.

Posttest Percentage - Pretest Percentage			
Age	Student age (in years)	29.66	9.53
Female	1 = Female; 0 = Male	.76	.43
Social Science	1 = Student major reported as anthropology, economics, political science, psychology, sociology, or unclassified social science; 0 = Professional Studies major	.68	.47
Prior Knowledge	Arithmetic average of the percentage of correct answers (out of 30 items) on the pretest.	43.89	11.13
Night Class	1 = Class began at 5 p.m. or thereafter; 0 = Class began prior to 5 p.m.	.12	.33
Class Size	Number of students in enrolled in course	30.83	4.10
Exam Mean	Arithmetic average of three examinations. Coded on a 0-100 point scale	80.63	12.72
Quiz Mean	Arithmetic average of twelve quizzes. Coded on a 0-10 point scale	8.25	1.15
Group Projects	Arithmetic average of the combined grades of Group Project 1 and Group Project 2. Coded into 12 descending numeric categories representing A to F, e.g., 12 = A; 11 = A-; 10 = B+, etc.	8.66	2.27

Note: Professional Studies majors, e.g., Early Childhood Education and Public Administration, serve as the omitted reference category for the academic major dummy variable.

Analytic Procedure

In order to identify student and course characteristics associated with learning, it first had to be established that knowledge was gained. The study's design generated appropriate data, while a statistical test determined if there were significant differences (i.e., learning) between pretest and posttest scores (Improvement, i.e., the dependent variable). A paired-sample *t* test was applied to each of the six sections of *Statistical Analysis*.

Hierarchical regression analysis is a technique in which independent variables are entered into an equation sequentially. Noting the increase in r-square due to particular independent variables, partitions the proportion of variance in the dependent variable accounted for by all the independent variables (Schutz et al., 1998). Hierarchical regression was used to: 1) evaluate the net effect of student characteristics (e.g., age, gender, prior knowledge) on their pretest-posttest difference and 2) assess the net effect of course characteristics (e.g., exams, group projects) on student's pretest-posttest improvement percentage. As such, in this study, the question is "How much of the total variance in students' learning (Improvement) is explained by specific independent variables, after controlling for the effects of all other independent variables?" Standardized regression coefficients represent the relative effect of each independent variable on the dependent variable.

Results

Pretest-Posttest Differences

A paired-sample *t* test was applied to each of the six sections of *Statistical Analysis*. Pretest-Posttest means, standard deviations, and differences appear in Table 2. The results reveal statistically significant (differences) gains in knowledge for each section and all courses combined. In sum, the pretest-posttest instrument consistently documents statistical knowledge gain, i.e., student learning.

Table 2. Pretest-Posttest Means, Standard Deviations and Differences

<i>Section</i>	<i>n</i>	<i>Pretest</i>		<i>Posttest</i>		<i>Difference</i>	<i>t</i>	<i>df</i>
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
1	31	45.3	11.2	67.9	11.8	22.6***	10.0	30
2	31	40.1	12.9	62.6	14.6	22.5***	8.5	30
3	36	46.0	10.7	66.6	13.9	20.6***	10.2	35
4	33	45.9	10.4	60.2	12.1	14.3***	8.2	32
5	32	43.3	11.5	64.0	12.5	20.7***	10.4	31
6	22	42.8	9.8	60.0	11.2	17.2***	7.9	21
1 – 6	185	44.0	11.2	63.8	13.0	19.8***	22.0	184

NOTE: The values for the difference column are the changes in the percentage correct from the pretest to the posttest. *** $p < .001$ (two-tail test)

Regression of Improvement Percentage on Student Demographics

By comparing the regression coefficients for age, female, and major in four different equations, change can be observed in the effects of these student characteristics (on learning), while controlling for prior knowledge and other independent variables. Table 3 displays results for Equation 1, Equation 2, Equation 3, and Equation 4. In the first equation, the dependent variable, Improvement, is regressed on age, female, and social science major. The estimated coefficient ($b = -3.58$) for Female is negative and statistically significant and can be interpreted as follows: Holding constant the effects of age and major, female students' improvement (knowledge gain) on the posttest is 3.58% less than male students. For Equation 1, *r*-square equals .020. This indicates that age, gender and major explain about 2 percent of the variation in the dependent variable, i.e., Improvement.

Table 3. Student Characteristics Regressed on Improvement (Equation 1), Controlling for Interaction between Age X Gender (Equation 2), Prior Knowledge (Equation 3) and Course Variables (Equation 4)

<i>Variable</i>	<i>Equation 1</i>		<i>Equation 2</i>		<i>Equation 3</i>		<i>Equation 4</i>	
	<i>b</i>	β	<i>b</i>	β	<i>b</i>	β	<i>b</i>	β
Age	.041 (.089)	.033	.300 ⁺ (.199)	.248	.154 (.183)	.127	.510*** (.183)	.422
Female	-3.581* (1.981)	-.133	5.788 (6.705)	.215	2.824 (6.126)	.105	8.838 (6.263)	.340

Social Science	1.192 (1.819)	.048	1.335 (1.816)	.054	2.454 (1.664)	.100	4.141*** (1.477)	.168
Female X Age			-.325 ⁺ (.222)	-.435	-.186 (.204)	-.249	-.479** (.208)	-.653
Prior Knowledge					-.435*** (.070)	-.419	-.680*** (.068)	-.651
Night Class							-6.160 (5.123)	-.180
Class Size							.003 (.414)	.001
Quiz Mean							.935 (.942)	.095
Exam Mean							.337*** (.086)	.382
Group Projects							.741** (.345)	.141
Intercept	22.774***		15.420***		38.300***		0.879	
R-Square	.020		.031		.201		.423	

NOTES: N = 185; b = unstandardized regression coefficient with standard errors in parentheses; β = standardized regression coefficient.

*** $p < .01$ (two-tail test); ** $p < .05$ (two-tail); * $p < .05$ (one-tail test); + $p < .10$ (one-tail)

Before concluding women acquire less statistical knowledge than men, an interaction term (Female X Age) is added to Equation 2. The regression coefficient ($b = -.325$) for Female X Age falls just short of statistical significance ($p < .10$). However, including the interaction, renders the coefficient ($b = 5.79$) for Female to a level that is no longer significant, while at the same time, Age ($b = .300$) approaches statistical significance (albeit at $p < .10$). R-square increases, from .020 in Equation 1 to .031 in Equation 2, an indication the interaction term explains some of the variation in Improvement percentage.

Regression of Improvement Percentage on Student Demographics and Prior Knowledge

To control for prior course knowledge, Equation 3 includes, as an independent variable, the percentage of correct items students attained on the pretest. The estimated coefficient ($b = -.435$) for Prior Knowledge exerts a negative and statistically significant effect on Improvement. Once again Age, Female and the interaction term, Female X Age, are not statistically significant. Notably, Equation 3 produces an r-square of .201, a substantial increase (nearly 7 times larger) over Equation 2, and an indication that Prior Knowledge explains most of the variation in the dependent variable. Holding constant all other independent variables, a one-percent increase in students' pretest score produces a .435% decrease in Improvement percentage. In other words, students improve less on the posttest if they entered the course knowing more (performed better on the pretest) than their peers.

Regression of Improvement on Student, Prior Knowledge, and Course Characteristics

Equation 4 includes several course characteristics (time of day, class size, and quiz, examination, and group project performance) absent from the previous equations. The new independent variables substantially increase the model's explanatory power. R-square (.423) in Equation 4 is more than twice the size generated in Equation 3 ($r^2 = .201$), meaning, course characteristics are a major predictor of the dependent variable (Improvement), over and above the effects of student demographics and prior knowledge.

Once again, Female ($b = 8.84$) is not statistically significant, however, both Age ($b = .510$) and the interaction term, Female X Age ($b = -.479$), attain significance in Equation 4. This can be interpreted as follows: The coefficient for Age ($b = .510$) represents the effect of age when Female equals 0, that is, when the student is male. Therefore, among men, each additional one-year increase in age produces (on average) a .510% gain in Improvement. The coefficient for the interaction, Female X Age ($b = -.479$), is the additional effect of age when the student is female, so the effect for women is $.510 - .479 = .031$. Therefore, each additional year increase in age for females predicts a rise of only .031% in Improvement. In sum, age has a statistically significant positive effect on learning for men, but little or no effect for women.

In Equation 4, the estimated coefficient ($b = 4.141$) for Social Science majors produces a positive and statistically significant effect on Improvement. This finding indicates, all other things being equal; Social Science majors score 4.14% higher on the dependent variable, than do Professional Studies students. Meanwhile, the effect of Prior Knowledge ($b = -.680$) remains negative and statistically significant.

Both examination and group project performance are positive and statistically significant predictors of learning. Specifically, holding constant the effects of all other independent variables, Equation 4 predicts that with each one-point increase in examination scores (Exam Mean), Improvement rises, on average, by .337%. Likewise, each additional grade category increase in Group Projects ($b = .741$) produces a .741% gain in knowledge.

Lastly, I compare standardized coefficients (β) to determine which independent variables have the greatest impact on Improvement, i.e., learning. The interaction term (Female X Age) and Prior Knowledge have the largest effect sizes ($\beta = -.653$ and $\beta = -.651$, respectively). These two variables explain the most variance in the dependent variable. Age ($\beta = .422$) is the third strongest predictor of Improvement, followed by Exam Mean ($\beta = .382$). Social Science major ($\beta = .168$) and Group Projects ($\beta = .141$) have the smallest effect sizes amongst the statistically significant independent variables.

Discussion

Based on significant paired-sample t tests, students' statistical knowledge is greater at the end of the semester than at the beginning of the course. This finding is important for teaching and assessment. Namely, a pretest-posttest design can document student learning or knowledge gained upon completion of university-level coursework.

Age, gender, major, and pretest scores combine to explain a large proportion (over 20 percent) of the variance in the dependent variable (Improvement). The results for the age x gender interaction require further discussion. Why does age have a much less positive effect on learning for women than men? One potential explanation is that as nontraditional (age) students, women may experience longer

interruptions between their college enrollments than do their male counterparts.³ If so, the advantages of life experience and maturity often attributed to nontraditional student success (Bye, Pushkar, & Conway, 2007; Carney-Crompton & Tan, 2002) may diminish with the knowledge loss associated with increasing time away from higher education. This may be especially salient for math-based courses, such as statistics.

Social science majors, compared to those in professional studies, gained more statistical knowledge. This finding cannot be attributed to social science students learning more because they entered the course knowing less. Pretest performance reveals social science majors (Mean = 45.3) scored higher than professional studies students (Mean = 42.5) on Prior Knowledge. One possibility is that social science majors have more prior exposure to statistical content in previous coursework (Bridges et al., 1998). This may reduce their anxiety and predispose them toward an interest in statistical analysis that leads to higher pretest and posttest scores (i.e., learning).

Among student characteristics, pretest scores (Prior Knowledge) account for the largest proportion (about 17 percent) of variation in Improvement, while demographic variables explain a little more than three percent. Overall, nearly half of the variance in learning is predicted by student background and prior knowledge. Students who scored low on the pretest, improved the most on the posttest, and acquired the most knowledge of statistical analysis. In sum, students learned more if they entered the course knowing less.

Course characteristics explain more than 50 percent of the variation in student learning. Time of day, class size, and quiz scores were not statistically associated with Improvement. However, examination and group project performance were significant positive predictors of statistical knowledge gain. These results are consistent with research in cognitive psychology that suggest examinations are learning devices that make course content more recallable than other activities that pre-expose students to material on which they will be evaluated (Little, et al., 2012; Little and Bjork, 2010; Little and Bjork, 2011). Examinations increase students' subsequent learning of course information by making ensuing study more effective (Little and Bjork, 2010; Little and Bjork, 2011), and the more time students spend studying, the more they learn (Arum and Roksa, 2011). The positive effect of group projects on learning is noteworthy. Students that earned high grades on groupwork, increased their knowledge of statistics more than students receiving lower grades. This suggests working collaboratively motivates students (though group interaction) to learn material related to content assessed on the posttest (McKinney and Graham-Buxton, 1993; Rau and Heyl, 1990; Yamarik 2007).

As indicated by comparison of standardized regression coefficients ($\beta = .141$ versus $\beta = .382$), group projects have an effect size less than half that attributed to students' performance on examinations. It is a challenge to explain this difference in view of the array of factors (many of which were not controlled in the present study) that can affect student learning. Some reasons to consider, include group differences in student ability, motivation, statistical knowledge, or group free riders. The latter explanation is supported by faculty observations and student comments. While working on their group projects, a few students complained to the professor about members not fulfilling their responsibilities, i.e., free riders. Therefore, some groups may have been more self-selective than others. Maybe the most conscientious students avoided free riders and found similarly motivated classmates. As a result, those groups received higher grades than groups composed of less committed students, and as individuals, the more conscientious students demonstrated more learning at the end of the course. Consequently, the positive effects of group project performance on statistical knowledge gain, may reflect the formation of homogeneous groups.

³ An independent *t*-test comparing the mean ages of women (30.1) and men (29.1) revealed no statistically significant differences.

Pedagogical Implications

Pretest-posttest assessment, once put into practice (and evaluated), can be used to improve teaching effectiveness. For example, posttest content on which students performed poorly can receive greater attention and class time when the course is next taught. Likewise, a pretest can detect the knowledge loss associated with extended time between mathematics coursework, such as that experienced by nontraditional students. This information may alert faculty of the need to review and/or refer such students for remediation services. Finally, pretests can also identify students with prior knowledge of course content, enabling faculty to devote less time to those areas in the future.

While the results of this study provide support for collaborative learning strategies, it also suggests some modifications to groupwork. First, faculty should make it difficult for students to free ride. For example, McKinney and Graham-Buxton (1993) recommend averaging individual and group grades on projects. Instructors could insist each student contribute at least one section to the project and require a table of contents identifying individual work. This would enable faculty to assign each student a grade for his or her contribution as well as awarding them a collective group grade. Second, professors might consider establishing permanent groups at the beginning of the semester. The groups could be formed voluntarily, assigned randomly, or based on ability (e.g., determined by a pretest or students' background in mathematics) in which students are placed in either mixed- or similar-ability groups (Borresen, 1990; Cumming 1983). These project modifications would provide faculty an opportunity to compare the relative effects of different group conditions on learning outcomes.

Conclusion

The results reported in this study suggest faculty take some precautions before committing to teaching strategies that have not been empirically associated with student learning. Nevertheless, the data are by no means representative of all institutions of higher education, and the conclusions drawn are best viewed as tentative. Therefore, I suggest the following areas for future investigation.

First, research is needed to identify course characteristics that improve learning at different types of institutions and on diverse student populations. Modifications in course design and implementation may be required for the effective application of instructional innovation in different environments. Second, more studies are required that connect pedagogical practices to actual student learning (i.e., direct assessment). Using student evaluations of teaching, attitude surveys, and even course grade point averages as learning outcomes do not adequately measure whether a particular technique increased students' skills and knowledge. Faculty should consider using gains in information content-learning to evaluate course outcomes (Gelles, 1980). Third, there is a need for more quasi-experimental assessments of pedagogy in college courses. This would include research designs that employ a systematic method of comparison, utilizing both pretest-posttest and experimental and control groups (Baker, 1985; Chin, 2002). For example, faculty assigned to teach multiple sections of statistics might use a "new" method of instruction in one section and compare the amount learned with a traditionally taught course.

This study has implications for higher education in the areas of pedagogy, student learning, and assessment. The results should interest faculty, in general, and those teaching statistics, one of the most challenging courses in the undergraduate curriculum, in particular. Investigation of teaching effectiveness and student learning are important, both for basic scholarship and because accountability studies by the federal government, state legislatures, and accrediting agencies have become increasingly outcomes based. Faculty seeking new ways to teach should continue to experiment with their pedagogy. At the same time, they must also systematically assess learning outcomes and be prepared

to make modifications in the application of such techniques when empirical evidence of student learning is not linked to the instructional activity.

Appendix

Appendix 1. Statistical Analysis Course Pretest and Posttest Content Areas.

Statistical Analysis Course Pretest and Posttest Content Areas

<i>Item#</i>	<i>Topic</i>
1.	Organizing Raw Data - Descriptive Statistics
2.	Frequency Distributions - Descriptive Statistics
3.	Contingency (Cross-Tabulation) Tables - Descriptive Statistics
4.	Contingency (Cross-Tabulation) Tables - Descriptive Statistics
5.	Histogram - Descriptive Statistics
6.	Scatter plot - Descriptive Statistics
7.	Skewness - Descriptive Statistics
8.	Percentiles - Descriptive Statistics
9.	Central Tendency (Mean) - Descriptive Statistics
10.	Central Tendency (Mode) - Descriptive Statistics
11.	Central Tendency (Median) - Descriptive Statistics
12.	Variance and Standard Deviation - Descriptive Statistics
13.	Normal Curve - Inferential Statistics
14.	Normal Curve - Inferential Statistics
15.	Confidence Interval - Inferential Statistics
16.	Hypothesis Testing - Inferential Statistics
17.	Hypothesis Testing - Inferential Statistics
18.	Hypothesis Testing - Inferential Statistics
19.	Hypothesis Testing - Inferential Statistics
20.	Hypothesis Testing - Inferential Statistics
21.	<i>t</i> Test - Inferential Statistics
22.	<i>t</i> Test - Inferential Statistics
23.	Analysis of Variance (ANOVA) - Inferential Statistics
24.	Analysis of Variance (ANOVA) - Inferential Statistics
25.	Chi-Square - Inferential Statistics
26.	Correlation
27.	Correlation
28.	Regression
29.	Regression
30.	Multiple Regression

References

- Arum, R. and Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago, IL: University of Chicago Press.
- Astin, A. (1993). *What matters in college? Four critical years revisited*. San Francisco, CA: Jossey-Bass.
- Auster, C. J. (2000). Probability sampling and inferential statistics: An interactive exercise using M&M's. *Teaching Sociology*, 28(4), 379-385.
- Baker, P. (1985). Does the sociology of teaching inform *Teaching Sociology*? *Teaching Sociology*, 12(3), 361-375.
- Bandalos, D. L., Finney, S. J., & Geske, J. A. (2003). A model of statistics performance based on achievement goal theory. *Journal of Educational Psychology*, 95(3), 604-616.
- Bjork, E. L., & Bjork, R. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M.A. Gernsbacher, R.W. Pew, L.M. Hough, & J.R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating contributions to society* (pp. 56-64). New York, NY: Worth Publishers.
- Blalock, H. M., Jr. (1987). Some general goals in teaching statistics. *Teaching Sociology*, 15(2), 164-172.
- Borresen, R. C. (1990). Success in introductory statistics with small Groups. *College Teaching*, 38(1), 26-28.
- Bridges, G. S., Gillmore, G. M., Pershing, J. L., & Bates, K. A. (1998). Teaching quantitative research methods: A quasi-experimental analysis. *Teaching Sociology*, 26(1), 14-24.
- Bye, D., Pushkar, D., & Conway, M. (2007). Motivation, interest, and positive affect in traditional and nontraditional undergraduate students. *Adult Education Quarterly*, 57(2), 141-158.
- Carney-Crompton, S., & Tan, J. (2002). Support systems, psychological functioning, and academic performance of nontraditional female students. *Adult Education Quarterly*, 52(2), 140-154.
- Center for Postsecondary Research. (2015). *Carnegie classification of institutions of higher education*. Indiana University School of Education. Bloomington, IN. <http://carnegieclassifications.iu.edu/index.php>
- Cerrito, P. B. (1999). Teaching statistical literacy. *College Teaching*, 47(1), 9-13.
- Chin, J. (2002). Is there a scholarship of teaching and learning in *Teaching Sociology*? A look at papers from 1984 to 1999. *Teaching Sociology*, 30(1), 53-62.
- DeCesare, M. (2007). Statistics anxiety among sociology majors: A first diagnosis and some treatment options. *Teaching Sociology*, 35(4), 360-367.

- Delucchi, M. (2006). The efficacy of collaborative learning groups in an undergraduate statistics course. *College Teaching*, 54(2), 244-248.
- Delucchi, M. (2007). Assessing the impact of group projects on examination performance in social statistics. *Teaching in Higher Education*, 12(4), 447-460.
- Fischer, H. W., III. (1996). Teaching statistics from the user's perspective. *Teaching Sociology*, 24(2), 225-230.
- Fisher-Giorlando, M. (1992). Sampling in a suitcase: Multistage cluster sampling made easy. *Teaching Sociology*, 20(4), 285-287.
- Forte, J. (1995). Teaching statistics without sadistics. *Journal of Social Work Education*, 31(2), 204-219.
- Garfield, J., & Chance, B. (2000). Assessment in statistics education: Issues and challenges. *Mathematical Thinking and Learning*, 2(1/2), 99-125.
- Gelles, R. J. (1980). Teaching sociology on teaching sociology. *Teaching Sociology*, 8(1), 3-20.
- Healey, J. F. (2015). *Statistics: A tool for social research*. 10th Edition. Cengage Learning.
- Helmericks, S. (1993). Collaborative testing in social statistics: Toward gemeinstat. *Teaching Sociology*, 21(3), 287-297.
- Little, J. L., & Bjork, E. L. (2010). Multiple-choice testing can improve the retention of non-tested related information. In S. Ohisson & R. Catrabone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1535-1540). Austin, TX: Cognitive Science Society.
- Little, J. L., & Bjork, E. L. (2011). Pretesting with multiple-choice questions facilitates learning. In L. Carlson, C. Holscher, and T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 294-296). Austin, TX: Cognitive Science Society.
- Little, J. L., Bjork, E. L., Bjork, R. A., & Angello, G. (2011). Multiple-Choice tests exonerated, at least of some charges: Fostering test-induced learning and avoiding test-induced forgetting. *Psychological Science*, 18(11), 1337-1344.
- Lomax, R. G., & Moosavi, S. A. (2002). Using humor to teach statistics: Must they be orthogonal? *Understanding Statistics*, 1(2), 113-130.
- Lucal, B., Albers, C., with Ballantine, J., Burmeister-May, J., Chin, J., Dettmer, S., & Larson, S. (2003). Faculty assessment and the scholarship of teaching and learning: knowledge available/knowledge needed. *Teaching Sociology*, 31(2), 146-161.
- Luce, C., & Kirnan, J. P. (2016). Using indirect vs. direct measures in the summative assessment of student learning in higher education. *Journal of the Scholarship of Teaching and Learning*, 16(4), 75-91. doi:10.14434/josotl.v16i4.193711.

- Macheski, G. E., Buhrmann, J., Lowney, K. S., & Bush, M. E. (2008). Overcoming student disengagement and anxiety in theory, methods, and statistics courses by building a community of learners. *Teaching Sociology, 36*(1), 42-48.
- Marson, S. M. (2007). Three empirical strategies for teaching statistics. *Journal of Teaching in Social Work, 27*(3/4), 199-213.
- McKinney, K., & Graham-Buxton, M. (1993). The use of collaborative learning groups in the large class: Is it possible? *Teaching Sociology, 21*(4), 403-408.
- Neuman, L. W. (1989). Which students learn the most, and why? A replication and extension of the Szafran pretest study. *Teaching Sociology, 17*(1), 19-27.
- Perkins, D. V., & Saris, R. N. (2001). A 'Jigsaw Classroom' technique for undergraduate statistics courses. *Teaching of Psychology, 28*(2), 111-113.
- Potter, A. M. (1995). Statistics for sociologists: Teaching techniques that work. *Teaching Sociology, 23*(3), 259-263.
- Price, B. A., & Randall, C. H. (2008). Assessing learning outcomes in quantitative courses: Using embedded questions for direct assessment. *Journal of Education for Business, 83*(5), 288-294.
- Rau, W., & Heyl, B. S. (1990). Humanizing the college classroom: Collaborative learning and social organization among students. *Teaching Sociology, 18*(2), 141-155.
- Schacht, S. P., & Stewart, B. J. (1990). What's funny about statistics? A technique for reducing student anxiety. *Teaching Sociology, 18*(1), 52-56.
- Schacht, S. P., & Stewart, B. J. (1992). Interactive/User-Friendly gimmicks for teaching statistics. *Teaching Sociology, 20*(4), 329-332.
- Schumm, W. R., Webb, F. J., Castelo, C. S., Akagi, C. G., Jensen, E. J., Ditto, R. M., Spencer-Carver, E., & Brown, B. F. (2002). Enhancing learning in statistics classes through the use of concrete historical examples: The space shuttle Challenger, Pearl Harbor, and the RMS Titanic. *Teaching Sociology, 30*(3), 361-375.
- Schutz, P. A., Drogosz, L. M., White, V. E., & Distefano, C. (1998). Prior knowledge, attitude, and strategy use in an introduction to statistics course. *Learning & Individual Differences, 10*(4), 291-308.
- Smith, B. (2003). Using and evaluating resampling simulations in SPSS and EXCEL. *Teaching Sociology, 31*(3), 276-287.
- Stork, D. (2003). Teaching statistics with study survey data: A pedagogical innovation in support of student learning. *Journal of Education for Business, 78*(6), 335-339.
- Strangfeld, J. A. (2013). Promoting active learning: Student-Led data gathering in undergraduate statistics. *Teaching Sociology, 41*(2), 199-206.

- Wagenaar, T. C. (2002). Outcomes assessment in sociology: Prevalence and impact. *Teaching Sociology, 30*(4), 403-413.
- Weiss, G. L. (2002). The current status of assessment in sociology departments. *Teaching Sociology, 30*(4), 391-402.
- Wilder, E. I. (2010). A qualitative assessment of efforts to integrate data analysis throughout the sociology curriculum: Feedback from students, faculty, and alumni. *Teaching Sociology, 38*(3), 226-246.
- Wybraniec, J., & Wilmoth, J. (1999). Teaching students inferential statistics: A "Tail" of three distributions. *Teaching Sociology, 27*(1), 74-80.
- Yamarik, S. (2007). Does cooperative learning improve student learning outcomes? *Journal of Economic Education, 38*(Summer), 259-277.