

2010

Are Statistics Labs Worth the Effort?--Comparison of Introductory Statistics Courses Using Different Teaching Methods

Jose H. Guardiola

Department of Mathematics and Statistics, Texas A & M University - Corpus Christi,

jose.guardiola@tamucc.edu

Nadina Duran-Hutchings

Department of Mathematics and Statistics, Texas A&M University - Corpus Christi (deceased),

Nadina.Hutchings@tamucc.edu

Hassan Elsalloukh

Department of Mathematics and Statistics, University of Arkansas at Little Rock, hxelsalloukh@ualr.edu

Follow this and additional works at: <https://scholarcommons.usf.edu/numeracy>



Part of the [Mathematics Commons](#), and the [Science and Mathematics Education Commons](#)

Recommended Citation

Guardiola, Jose H., Nadina Duran-Hutchings, and Hassan Elsalloukh. "Are Statistics Labs Worth the Effort?--Comparison of Introductory Statistics Courses Using Different Teaching Methods." *Numeracy* 3, Iss. 1 (2010): Article 5. DOI: <http://dx.doi.org/10.5038/1936-4660.3.1.5>

Authors retain copyright of their material under a [Creative Commons Non-Commercial Attribution 4.0 License](#).

Are Statistics Labs Worth the Effort?—Comparison of Introductory Statistics Courses Using Different Teaching Methods

Abstract

This paper compares the academic performance of students in three similar elementary statistics courses taught by the same instructor, but with the lab component differing among the three. One course is traditionally taught without a lab component; the second with a lab component using scenarios and an extensive use of technology, but without explicit coordination between lab and lecture; and the third using a lab component with an extensive use of technology that carefully coordinates the lab with the lecture. Extensive use of technology means, in this context, using Minitab software in the lab section, doing homework and quizzes using MyMathlab ©, and emphasizing interpretation of computer output during lectures. Initially, an online instrument based on Gardner's multiple intelligences theory, is given to students to try to identify students' learning styles and intelligence types as covariates. An analysis of covariance is performed in order to compare differences in achievement. In this study there is no attempt to measure difference in student performance across the different treatments. The purpose of this study is to find indications of associations among variables that support the claim that statistics labs could be associated with superior academic achievement in one of these three instructional environments. Also, this study tries to identify individual student characteristics that could be associated with superior academic performance. This study did not find evidence of any individual student characteristics that could be associated with superior achievement. The response variable was computed as percentage of correct answers for the three exams during the semester added together. The results of this study indicate a significant difference across these three different instructional methods, showing significantly higher mean scores for the response variable on students taking the lab component that was carefully coordinated with lecture. The authors acknowledge limitations of this study; however, their findings are in agreement with their classroom experience.

Keywords

instructional methods, teaching methods, student achievement, introductory statistics, statistics education

Creative Commons License



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 License](https://creativecommons.org/licenses/by-nc/4.0/)

Introduction

Advancement of technology has generated opportunities and challenges for teaching introductory statistics and other courses that contribute to quantitative literacy (QL) across the curriculum. The Guidelines for Assessment and Instruction in Statistics Education (GAISE 2005) recommend, among other things, the use of technology for developing conceptual understanding and analyzing data. In a traditional classroom setting, however, students have limited or no access to computers with specialized statistical software. With GAISE in mind, faculty at Texas A&M University-Corpus Christi (TAMUCC) developed a lab component for an introductory statistics course that promotes the use of technology and active learning with problems based on scenarios. We believe our experience and what we have learned from it will be of interest to others in the QL community who are considering adding computer labs to their courses.

TAMUCC offers a “traditional” three-hour statistics lecture course, MATH 2342. The university is in the process of implementing a new four-hour course, MATH 1442, which includes a lab component of two hours per week and promotes a more intensive use of technology. This new course, MATH 1442, has been added as a core-curriculum option and is expected to eventually replace the traditional course, MATH 2342. Even though the two courses, MATH 1442 and MATH 2342, have different course numbers, their academic contents are essentially the same. They have a combined enrollment of 250 to 300 students every semester.

Active learning has been the focus of many published studies. Problem-based learning (PBL) in statistics courses stresses the connection between statistical reasoning and scientific inquiry as suggested by Boyle (1999). Only a few studies have compared teaching of statistics under the absence or presence of technology to any extent. Utts et al. (2003) compared Web-based instruction with traditional teaching and concluded that the performance of students in the Web-based instruction equaled that of the traditional students, but they were slightly less positive in their subjective evaluation of the course. Basturk (2005) suggested that computer-aided instruction enhances the educational experience when it is used as a supplement to the regular lecture in teaching introductory statistics courses. In addition, a very complete review of the literature at that time regarding simulation methods for teaching statistics is available in the article by Mills (2002). In particular, some computer-simulation techniques such as the bootstrap may add value in teaching statistical concepts, specifically sampling distributions (see Hesterberg, 1998, for details). Also, Velleman (2000) discussed some design principles aimed at developing multimedia-based statistics material including animation and student interaction with computer graphic applets. In our review of the literature, however, we have not been able to find any published studies that

focus on the comparison of teaching statistics with a lab component using relevant technology versus the traditional, lecture approach with no lab component.

The Problem

It is important for the Department of Mathematics and Statistics and for our university to determine whether the learning outcomes are better met for students taking this new course when compared with students in a traditional lecture-based class. Our question can be stated as:

Is there a difference in academic achievement for students taking an elementary statistics course with an integrated lab component, compared with the achievement of students taking a more traditional elementary statistics course without an integrated lab component?

A positive answer to this question would help to justify the additional resources dedicated to the new lab component and the careful coordination between the course instructor and lab instructor.

The purpose of the study reported here is two-fold: (1) to evaluate the effectiveness of including an integrated lab component in an elementary statistics course and (2) to identify particular individual student characteristics that may contribute to superior achievement with any of the three instructional methods mentioned in the next section.

Methodology

Design

Three sections of MATH 1442 and one section of MATH 2342 were offered in the spring semester, 2007. During summer, 2007, one section of each was taught. For the traditional course, MATH 2342, homework was assigned regularly and was graded by hand; eight quizzes were given during class time for every major topic throughout the semester; and two midterm exams and a comprehensive final examination were administered. For the newly implemented course, MATH 1442, a similar amount of homework was assigned, with similar content, but the exercises required the use of MyMathLab¹. Homework was graded immediately by MyMathLab throughout the semester in the new course. The lab component for MATH 1442 included exercises in the form of scenarios from the lab manual

¹ MyMathLab, © 2009 by Pearson Education, Inc. CourseCompass is a trademark of Pearson Education, for more details see www.coursecompass.com (accessed Dec. 10, 2009).

developed by Champine (2006). As in MATH 2342, two midterm exams and a comprehensive final examination were administered in MATH 1442.

The scenarios for the lab component of MATH 1442 are hypothetical problems designed to motivate the student to apply previously learned concepts during lecture. They are examples of problem-based learning (PBL) and promote learning by getting students to apply statistical theory to solve practically oriented problems. Appendix A lists the scenarios, and Appendix B gives a detailed example of two of them.

Our study compares the performance of 118 students during the spring and summer sessions of 2007, divided into four separate elementary statistics sections, taught by the same instructor, utilizing three different instructional methods. Students from 51 different majors were enrolled in these courses, from all colleges on campus. There were three methodologies: a control and two treatments.

- *Control.* MATH 2342 has never included a lab component and is traditionally taught without extensive use of technology. MATH 2342 was taught during spring 2007 and summer 2007. No statistical package was used for assignments. The textbook used was Bluman A., *Elementary Statistics*, 4th Ed., 2006.
- *Treatment 1.* The second course methodology is introduced into a section of MATH 1442 by including a lab component that uses scenarios and promotes an extensive use of relevant technology, with assignments and scenarios designed accordingly. Although lab topics and lecture topics were chosen to be complementary, each component was taught without explicit reference to the other. Scenarios were discussed and assigned as lab homework with no explicit reference to the lecture. The statistical package Minitab was used to perform statistical analysis as part of the lab activities. This course was taught during spring 2007. The textbook used was Triola, M., *Elementary Statistics*, 10th Ed., 2006.
- *Treatment 2.* The third course methodology was taught in sections of MATH 1442 by using a lab component and relevant technology, but also carefully coordinating labs with lecture content. Scenarios were discussed during lab time and were adapted and scheduled in order to follow closely the contents of previous lecture. The scenario seen during lab time was also discussed during lecture. Minitab was used to perform statistical analysis as part of the lab activities. Problems discussed and solved by hand during lecture time were discussed at the lab and they were solved again using Minitab software. The textbook used was Triola, M., *Elementary Statistics*, 10th Ed, 2006.

An effort was made to keep all factors that may influence student outcomes as uniform as possible across all groups. Assigned homework differed across the methods treatments; this difference was considered part of the treatment, as the homework for methods Treatments 1 and 2 was more technology oriented. In addition, the textbooks used were different for the control and methods Treatments 1 and 2; however, they also were considered as part of the treatment. An effort was made to keep the same lecture content by using the same class examples and lecture style.

After approval was obtained from the Institutional Review Board at Texas A&M Corpus Christi, data were collected from students and from official university records, including grade point average (GPA), attendance, age, gender, academic major, and class standing. These data were used as covariates as they may contribute to explain the academic performance of students. Summary statistics for the first four of these covariates are shown on Table 1.

The variable “academic major” was replaced by the variable “college.” This change from a potential of 51 academic majors to only five colleges still clusters students by different quantitative preparations across colleges.

Table 1
Summary Statistics for Covariates

Covariates	GPA			Attendance		Age		Gender
	Treatment	N	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Control Spring	52	2.90	0.72	85	17	23	4.7	58
Control Summer	23	2.80	0.57	87	19	24	7.7	65
Treatment 1 Spring	26	2.79	0.65	73	28	22	3.6	69
Treatment 2 Summer	17	2.79	0.61	82	20	22	2.6	65
Total	118	2.84	0.66	82	21	23	5.0	63

Additionally, a multiple intelligences (MI) instrument available online² was administered to students. Gardner (1983) developed the theory that there exist seven different intelligences and learning styles. The MI instrument, based on the concepts from Gardner (1983), attempts to measure the seven types of learning style: verbal, mathematical/logical, musical, spatial, body/kinesthetic,

² More information on Multiple Intelligences (MI) and examples of the MI instrument can be found at <http://www.mitest.com> (Accessed on February 2007).

interpersonal, and intrapersonal. Each of these items was evaluated on a scale of 0–12, where a higher score indicates a higher ability related to that type of intelligence. Summary statistics for the multiple intelligence tests are shown in Table 2.

Table 2
Summary statistics for multiple intelligence test results.

Method	Verbal	Math Logical	Music	Spatial	Body Kinaes- thetic	Inter- personal	Intra- personal	Total
Control Spring	4.9 (2.2)	6.2 (2.5)	6.9 (2.8)	5.5 (2.1)	7.9 (2.2)	6.5 (2.5)	5.2 (1.9)	43.2 (9.2)
Control Summer	5.3 (1.7)	7.2 (3.2)	6.5 (3.0)	6.7 (2.5)	7.5 (2.5)	6.0 (2.1)	6.0 (1.8)	45.2 (10.0)
Treatment 1	4.8 (2.1)	5.5 (2.4)	6.6 (1.9)	5.3 (3.5)	6.9 (2.3)	5.9 (2.6)	5.9 (2.5)	40.8 (9.4)
Treatment 2	4.2 (2.8)	6.5 (3.1)	7.9 (3.3)	7.1 (2.7)	8.5 (2.3)	7.1 (1.6)	5.6 (2.1)	46.8 (11.2)
MIT Total	4.8 (2.1)	6.3 (2.7)	6.9 (2.7)	6.0 (2.7)	7.7 (2.3)	6.3 (2.3)	5.6 (2.0)	43.7 (9.8)

Notes: The number at top of each cell is the mean and the number in parentheses is the standard deviation. Sample sizes are shown in Table 1. MIT stands for multiple intelligence tests.

These measurements were intended to be used as covariates in order to explore whether differences in individual students' preferred learning styles may contribute to a superior performance under any of these three different teaching methods. Studies done by Snow (1989) suggest that different instructional methods are more or less effective depending upon particular individual aptitudes and learning styles. More recently, Denig (2004) has compared the theories of multiple intelligences and learning styles to suggest ways that teachers may be able to improve student learning.

The same or very similar midterm tests and final examinations were administered to the three treatments during spring and summer semesters, adjusting only for differences in class duration for different courses by inserting some additional questions for the longer classes. Those additional questions were ignored for the purpose of this study. Control and Treatment 1 were applied during the spring semester, and a new Control and Treatment 2 were applied during the summer session.

Exploring the Covariates

An initial exploratory analysis was undertaken for all variables used as covariates. These explanatory variables, including gender, GPA, college of academic major,

class standing and attendance were analyzed and compared across treatments in order to detect differences in student subsamples across the different treatments. No statistically significant differences were detected across the different treatments at the 0.05 significance level, except for attendance; however, it was not found statistically significant in the final reduced model.

Also, in order to assess uniformity between spring and summer sessions, comparisons between the Controls for spring and summer were made, and no significant differences were found among the students' multiple intelligences, or across GPA, attendance, age, college, class standing, gender and total exam scores (response variable). These analyses helped to lessen concerns about initial differences between those students who enrolled in the spring classes and those who enrolled during the summer.

Correlations among continuous covariates were explored, and the correlation matrix is shown in Table 3. Even though some correlations are significant there is no evidence of significant collinearity per the maximum value of the variance inflation factor (VIF), 1.66, which is considered low. The presence of a few outliers did not seem to have an effect on the overall results. Cook's distances were checked and all observations seem to be within reasonable ranges with no indication of highly influential points.

Table 3
Correlation Matrix for Continuous Variables.

	Total MIT	Age	GPA	Attendance	Final Scores
Total MIT	1.0000	0.0652	-0.0021	-0.0063	0.0530
Age	0.0652	1.0000	-0.2035*	-0.0540	-0.1350
GPA	-0.0021	-0.2035*	1.0000	0.3671*	0.6591*
Attendance	-0.0063	-0.0540	0.3671*	1.0000	0.2393*
Final Scores	0.0530	-0.1350	0.6591*	0.2393*	1.0000

*correlation is significant at the 0.05 level.

Analyzing the Response Variable

The response variable is defined as the sum of the total number of correct answers on each of the three exams added together and expressed as a percentage of the total number of questions (see Appendix C for a sample of selected questions on the final exam). The independent variables considered for our analysis are: each students' multiple intelligences scores for every type of intelligence, GPA, college of the academic major, age, class standing, gender and attendance.

A general linear model (GLM) was used to assess the outcomes from the three different methods while controlling for covariates. The full model considered all those interactions of the second order that were sensible to include (Ramsey and Schafer, 2002). The full model is described in detail in Appendix D.

A Tukey-Kramer test was used for multiple comparison tests after controlling for covariates. This procedure is recommended for unequal sample sizes (see sample sizes on Table 1); for more details see Kramer (1956), or Kirk (1995).

Starting with the full model described in Appendix D, a reduced model was found after variables were dropped from the full model by using backward elimination and minimizing the Schwartz Bayesian Criteria (SBC) (Ramsey and Schafer, 2002); for more details on the computer output see Appendix E.

Additionally, we performed a manual elimination starting from the full model. This time we dropped the variable with the largest p -value at each step until only variables significant at the 0.05 significance level were kept. Main effects were not dropped if there was an interaction present that included the main effect. This manual elimination process produced the same reduced model as the automatic backward elimination.

Forward elimination and stepwise elimination were attempted as well, and they produced slightly different models as there is no such thing as a “unique correct model” (Ramsey and Schafer 2002). In the interest of parsimony, we prefer the model found by backward elimination or manual elimination.

To determine the strength of association, partial omega-squared ω^2 was calculated. A simple estimate of effect size was then computed from omega-squared using $\hat{f} = \sqrt{\omega^2 / (1 - \omega^2)}$; for more details see Kirk (1995, pp. 397-399).

Results

Measurements obtained from the multiple intelligences test were found to be non-significant at the 0.05 significance level; therefore, all these variables were dropped from the full model during the elimination process as there is insufficient evidence to indicate that they contribute to the explanation of the final scores.

The final reduced model includes only significant variables at the 0.05 significance level. These are: treatment ($p = 0.003$), college ($p < 0.0001$) and GPA ($p < 0.0001$). Adjusted means, also called least squares or estimated marginal means, and their corresponding confidence intervals are shown in Table 4 for the three instructional methods.

The Tukey-Kramer multiple comparison indicated that there is a significant difference between Treatment 2 and Control ($p = 0.0065$). Students' exam scores are about eight percentage points higher for Treatment 2 than for those students under the Control after adjusting for covariates (Fig. 1). There is a significant

difference between Treatment 1 and Treatment 2 ($p = 0.0042$). However, no significant difference was found between Treatment 1 and Control ($p = 0.77$). For more details on the difference of least square means or adjusted means see Table 5.

Table 4
Adjusted Means for the Three Instructional Methods and the
Corresponding Confidence Intervals.

Method	Final LSMEAN	95% Confidence Limits
Control	60.5	(57.4, 63.5)
Treatment 1	59.0	(55.1, 62.8)
Treatment 2	68.1	(63.5, 72.7)

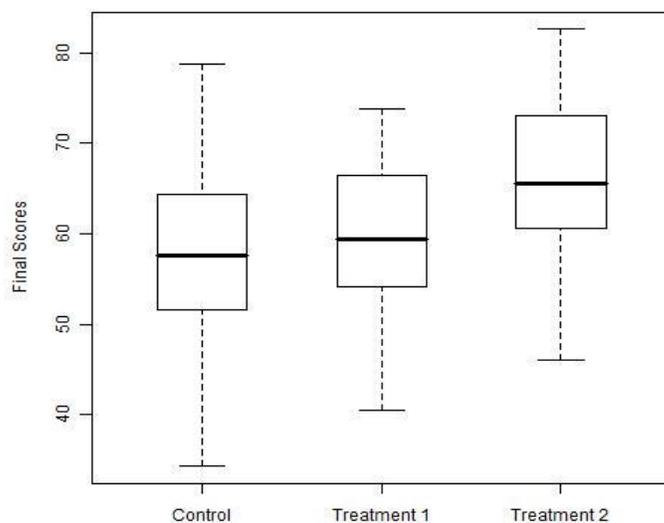


Figure 1: Boxplots of final scores adjusted for covariates for all treatment levels (Residuals were plotted after adding back adjusted means to each score). Control was administered during the spring and summer session. Treatment 1 was administered during the spring semester; Treatment 2 was administered during the summer semester.

Table 5
Difference between Adjusted Means (Least Squares Means)

Comparison	Mean Difference	95% Confidence Limits for Differences
Control vs. Treatment 1	1.5	(-3.6, 6.7)
Control vs. Treatment 2	-7.6	(-13.4, -1.8)
Treatment 1 vs. Treatment 2	-9.1	(-15.8, -2.5)

The variable “college” was used to capture a quantitative orientation of the student depending on college and it was found statistically significant at the final model. Using the Tukey-Kramer test we found non-significant differences among the College of Business, College of Science and Technology and the College of Nursing and Health Sciences, and significantly lower scores for the College of Education and College of Liberal Arts. See the computer output of Appendix E for further details.

After fitting the reduced model including only statistically significant terms (at the 0.05 significance level), the analysis of residuals showed a normal pattern consistent with the theory of general linear models.

The strength of association for treatments was computed to be $\omega^2 = 0.083$, which means that treatments can explain about 8% of the variation of final scores. The effect size was estimated to be $\hat{f} = 0.30$, which is between a small and medium effect.

Discussion

Regarding the statistically significant difference found between Treatment 1 and Treatment 2, it is important to notice that the extra hour of teaching is common to both treatments; therefore, it cannot be the cause of improvement for the latter. Similarly, it should be noticed that the integration of lab and lecture was the only difference between Treatment 1 and Treatment 2 as these labs and lecture contained essentially the same information. Therefore, the difference in performance for both treatments could not be attributed to more relevant information provided on one of the labs, as it is clearly not the case. Based on these differences, this study suggests that use of technology and additional time spent in the lab *in itself* is not enough to produce favorable outcomes in student academic performance. *This study does suggest that effectiveness of teaching a successful lab relies mainly on a careful coordination between lab activities and lecture topics.* Without such coordination, students may lose the connection between lab activities and the main concepts of the lecture. It is very important to use every possible opportunity to mention to the students the relationship between

lab activities and concepts studied during lecture. The lab activity must be seen as a complementary activity for reinforcing the theory examined during lecture. The lab instructor and the lecturer must coordinate their efforts by mentioning each other's activities and referring to the concepts taught during the other component in order to improve the educational experience for students in an introductory statistics course. It is important to highlight that, along with the mentioned supportive quantitative evidence, within the limitations mentioned below, these results are consistent with the authors' classroom experience.

The authors acknowledge some limitations of this study. Students were not randomly assigned to the treatments; however, at the time of this study it was not possible to overcome this limitation, as students choose their classes based on their perceived needs. There was no valid mechanism to randomly assign them into classes.

Another potentially confounding factor is the different duration of classes between spring courses (15 weeks) and summer courses (5 weeks). It appears to be non-significant, given that no statistical difference was found for the Control courses between spring and summer outcomes when comparing all covariates.

Different textbooks were used for Control and Treatments; however, they were considered part of the treatment, as one of the textbooks facilitates further the inclusion of technology. The authors cannot ignore that textbooks and homework could also be considered as confounding factors.

There are some other unmeasured factors that may affect the response variable, like motivation of the students and more subtle factors that were not included in this study. Additional questions included on the exams for Treatment 2 that were not part of the response variable may also be considered as a confounding factor; however, if there is any effect for these additional questions, it may strengthen our conclusion of better results for Treatment 2.

Different class sizes may also be a confounding factor as smaller classes facilitate student-professor interaction. Treatment 2 has the smallest class size. Sample sizes were not as large as preferred for the purpose of this study; however, they seem to be large enough to detect significant differences across teaching methods.

As discussed in the previous section, treatments explain about 8% of the variation on the dependent variable final scores which correspond to an effects size of between small and medium. Even though the mentioned associations were found to be statistically significant, the causes of such associations cannot be determined with certainty as is typical with these types of studies with numerous possible confounding factors.

Conclusions

We believe that the statistics labs are worth the additional time and effort dedicated to them provided that they are carefully coordinated with lecture topics. Also, we consider that these results support our classroom experience since students taking the labs coordinated with lecture seem to be more aware and inquisitive during lecture. Although we caution readers to keep in mind previously mentioned limitations before trying to extrapolate these results to their own particular environment and experience, we do recommend that QL educators who are considering integrating computer lab work into their QL courses give serious consideration to how they will coordinate the lab work with the other components of their courses.

Future Directions

The authors suggest improvements, should this study be repeated, such as randomly assigning classroom of students to different treatments in a larger setting. Larger sample sizes would be highly desirable. Group randomized designs using hierarchical models may also improve the current design.

Additionally, utilization of the same textbooks, same class lengths and class sizes, all treatments administered during the same semester, would be conducive to stronger conclusions by avoiding possible confounding factors previously mentioned. Similarly, including individual characteristics of the subjects such as motivation, aptitude and interest may also be enlightening.

Acknowledgments

The authors are thankful to the former Chair of the Department of Mathematics and Statistics, Blair Sterba Boatwright and current Chair George Tintera, for their support in the completion of this project, to Marilyn Spencer at the College of Business and Elaine Young at the Mathematics and Statistics Department both at Texas A&M University Corpus Christi, Thomas McMillan at the Department of Mathematics and Statistics at the University of Arkansas at Little Rock, Maria Guardiola and Doyle Boese for their help editing this article and to Joseph Champine for providing material for the labs. The authors are thankful to the reviewers and editor Len Vacher for their valuable suggestions in improving this paper.

References

- Basturk R. 2005. The effectiveness of computer-assisted instruction in teaching introductory statistics. *Educational Technology & Society*, 8 (2): 170–178.
- Boyle C. 1999. A problem-based learning approach to teaching biostatistics. *Journal of Statistics Education*. 7(1).
<http://www.amstat.org/publications/JSE/secure/v7n1/boyle.cfm> (accessed Nov. 1, 2009).
- Champine J. 2006. *Math 1442: Introductory statistics lab manual*. Texas A&M University – Corpus Christi.
- Denig S. 2004. Multiple intelligences and learning styles: Two complementary dimensions. *Teachers College Record*, 106(1): 96–111.
<http://www.tcrecord.org/content.asp?contentid=11513> (with membership; accessed July 12, 2009) .
- GAISE. 2005. *Guidelines for assessment and instruction in statistics education: College report*. American Statistical Association .
<http://www.amstat.org/education/gaise/GAISECollege.htm> (accessed Nov. 1, 2009).
- Gardner H. 1983. *Frames of mind: The theory of multiple intelligences*. New York, NY: Basic Books.
- Kirk R. 1995. *Experimental design*. 3rd. ed. Pacific Grove CA: Brooks/ Cole Publishing Company.
- Hesterberg, T. (1998), Simulation and bootstrapping for teaching statistics, *Proceedings of the Statistical Education Section*, American Statistical Association, 44–52.
- Kramer C. Y. 1956. Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics*, 12: 307–310. <http://dx.doi.org/10.2307/3001469>
- Mills, J. 2002. Using computer simulation methods to teach statistics: A review of the literature, *Journal of Statistics Education*, 10(1).
<http://www.amstat.org/publications/jse/v10n1/mills.html?ref=Guzels.TV> (accessed Nov. 1, 2009).
- Ramsey F., and D. Schafer. 2002. *The statistical sleuth*. 2nd Ed., Pacific Grove CA: Duxury.
- Snow Richard. 1989. Aptitude-treatment interaction as a framework for research on individual differences in learning. In *Learning and individual differences: Advances in theory and research*, ed. Ackerman, P, R. Sternberg and R. Glaser, 13-59. New York: W.H. Freeman/ Times Books/ Henry Holt, & Co.
- Utts J., B. Sommer, C. Acredolo, M. Maher, and H. Matthews. 2003. Study comparing traditional and hybrid Internet-based instruction in introductory statistics classes. *Journal of Statistics Education*: 11(3).
<http://www.amstat.org/publications/jse/v11n3/utts.html> (accessed Nov 1, 2009).
- Velleman, P. 2000. Design principles for technology-based statistics education. *Metrika*, 51(1): 91–104. <http://dx.doi.org/10.1007/PL00003977>