

## **Strength in Numbers: A Multidisciplinary, Project-based Course in Introductory Statistics**

Lisa Dierker<sup>1</sup>, Emmanuel Kaparakis, Jennifer Rose,  
Arielle Selya, and David Beveridge  
*Wesleyan University, Middletown, CT 06459*

---

### **Abstract**

This paper describes a multi-disciplinary, project-based course. The course is aimed at providing greater curricular access to applied statistics for students across both divisional and departmental boundaries and includes lecture and laboratory components as well as intensive individualized instructor and peer support. Learning materials and teaching strategies were designed to be structured enough to allow students to consistently move forward with their research projects, yet broad enough to encourage them to creatively and independently explore their questions by actively driving the decisions involved in inquiry. In this way, the support each student receives is dictated by their own research question and the results at each stage of their project. We describe the course around the six recommendations of the ASA-endorsed Guidelines for Assessment and Instruction in Statistics Education (GAISE) which foster opportunities for project-based work through an emphasis on the use of real world data, active learning, conceptual understanding rather than memorization, and the use of technology.

**Keywords:** Project-based; Data; Introductory statistics.

---

The days of “silo” science have ended—collaboration and inter-disciplinarity are now viewed as essential for solving the most important problems faced in the United States and the world (National Academies, 2004). The shift away from the solitary researcher to team research with scholars collaborating with others within and across disciplines has occurred in the natural sciences, computing sciences, engineering, the social sciences and the humanities (Wuchty, Jones, & Uzzi, 2007). Curricula that impart this kind of inclusive and flexible thinking and communication will best foster the development of students who will not only be able to engage in interdisciplinary scholarship, but will also be among the most scientifically literate citizens in our society.

To achieve this inter-disciplinarity, it is believed that project-based learning will play a central role. Project-based learning is most commonly defined as an instructional approach based on authentic, real-world activities that are aimed at engaging student interest and enthusiasm (BIE, 2012). Designed to answer a question or solve a problem, this approach allows students to face challenges that lead to answers, reflect on ideas and

---

<sup>1</sup> Corresponding author's email: [ldierker@wesleyan.edu](mailto:ldierker@wesleyan.edu)

make decisions that affect project outcomes. In fact, an emerging literature shows that project-based learning in many contexts is more effective in promoting deep thinking, the ability to apply knowledge, communication and reasoning skills when compared to traditional didactic approaches (e.g. Harada & Yoshina, 2004; Hickey, Wolfe, & Kindfield, 2000; Hickey et al., 1999; Langer, 2001; Lynch, Kuiper, Pyke, & Szesze, 2005).

Given that data is one of the most salient points of intersection among diverse disciplines, we describe in this paper a project-based curriculum that teaches diverse skills through the process of project-based statistical inquiry. Notably, statistics education has generally remained at the poles of discipline-specific instruction (e.g. psychology and economics) or is delivered more generically, without clear links to the work of any particular discipline. Thus, a central challenge is the development of curricula that not only serves diverse majors, but also sparks communication, reasoning and collaboration that clearly crosses the traditional disciplinary boundaries. This goal can best be achieved through project-based curriculum that allows students to “decompose their topic, identify key components; abstract and formulate different strategies for addressing it; connect the original question to the statistical framework; choose and apply methods; reconcile the limitations of the solution; and communicate findings” (Nolan & Temple Lang, 2009).

To begin to address this challenge, we have created a multi-disciplinary, project-based course in introductory statistics that can serve students from diverse majors. Learning materials and teaching strategies were designed to be structured enough to allow students to consistently move forward with their research projects, yet broad enough to encourage them to creatively and independently explore their questions by actively driving the decisions involved in statistical inquiry. In this way, the support each student receives is dictated by their own research question and results at each stage of their project.

### **Course Curriculum and Logistics**

This project-based course differs from standard introductory statistics courses in several ways. Where traditional statistics courses employ a building-block approach which covers relatively few statistical tools in a serial manner and in the absence of a context in which to apply them, this course entails learning statistics as students answer their own questions and choosing among many statistical tools presented in a parallel manner. This course promotes learning by creating a context in which students familiarize themselves with several possible data sets, formulate a statistical question, choose among several available statistical tools, apply an appropriate method, and communicate their findings. This process has not only given students grounding in basic statistics, but also helped them develop diverse skills related to inquiry beyond those that are exclusively statistical. The semester-long course met 3 to 4 times a week for a total of 4 hours. During the first offering of the course, substantive content was presented during lecture sessions with optional readings made available on-line (ratio of lecture to laboratory sessions was 2 to 1). For the second offering, lectures were stream-lined and the Carnegie Mellon Open Learning Initiative (OLI) was added as required online reading material and interactive support activities. The ratio of lecture to laboratory sessions was also reversed (i.e. 1 to 2). Table 1 shows a list of topics, along with project activities.

**Table 1. Weekly Topics and Activities.**

<b>Week</b>	<b>Topic</b>	<b>Activities</b>
1	Data, data sets and data documentation	Exploring data documentation
2	Reviewing the Literature	Literature review
3	Scientific writing and referencing	Draft Research Plan
4	Statistical software	Running a basic procedure and error checking data
	Review of software specific code for data management	Data management
5	Descriptive statistics and data visualization	Graphing means and frequency distributions
6	Inference	Graphing with two variables
7	Comparing means (ANOVA) and tests of categorical independence (Chi Square)	Testing and interpreting bivariate associations
	Post hoc tests	Testing and interpreting bivariate associations and post hoc comparisons
8	Correlation and Regression	Testing, graphing and interpreting bivariate associations
9	Multiple Regression	Replicating bivariate results with multiple regression
10	Logistic Regression	Replicating bivariate results with logistic regression
11	Confounding	Multivariate modeling
12	Statistical interactions	Continued multivariate modeling
13	Poster Presentation	Drafting research poster
14	Presenting Research Findings Orally	Drafting research poster and preparing oral presentations

On-line lecture clips are designed to provide students with adequate substantive and practical background for completing exams and cumulative laboratory assignments toward the completion of their research project. Each lecture focuses on a topic that allows students to make progress on their independent project. Laboratory instructors and teaching assistants are available during smaller lab sessions (20 to 25 students) to support students in completing their research project. Attendance is mandatory. Additional individualized instructional support from peer tutors is available 13 hours a day throughout the semester to a) minimize common frustrations experienced when working with data; and b) maximize productive project-based learning.

## Guidelines for Assessment and Instruction in Statistics Education

We discuss how the course described above relates to the current guidelines for statistics education approved by the American Statistical Association (see The Guidelines for Assessment and Instruction in Statistics Education – GAISE; amstat.org and (Franklin & Garfield, 2006) which foster opportunities for project-based work.

These six recommendations are:

1. Emphasize statistical literacy and develop statistical thinking;
2. Use real data;
3. Stress conceptual understanding rather than mere knowledge of procedures;
4. Foster active learning in the classroom;
5. Use technology for developing conceptual understanding and analyzing data;
6. Use assessments to improve and evaluate student learning

### *Emphasize statistical literacy and develop statistical thinking*

The course is designed around student research projects of their own choosing and offers intensive hands-on experience in not only using statistics, but also in some of the broader aspects of applied research, in which they can contextualize the statistical skills they are taught during class lectures. In the laboratory and in homework assignments, students develop skills in a) generating testable hypotheses; b) conducting a literature review; c) understanding the structure of large data sets; d) formatting and managing data; e) conducting descriptive and inferential analyses; and f) reporting and interpreting results. Laboratory instructors provide support by engaging students in discussions about their research topics and their experiences throughout the research process; providing statistical guidance and feedback; teaching statistical software syntax and de-bugging errors; and occasional brief lectures in support of the current statistical topics. Projects are presented at the end of the semester at a research poster session in which students have the opportunity to describe their process of inquiry, including the different decisions made along the way, their premises, conclusions and any barriers that they faced.

### *Use real data*

In the first week of class, students choose from a number of cutting-edge data sets representing several disciplines. Specifically, we provide data and supporting resources (e.g. data documentation, previously published scientific literature, grant applications describing the research, etc.) for studies that are made accessible by faculty and data that are publically accessible through national archives. Data sets for the course, are selected based on a) the completeness and clarity of their documentation; b) size in terms of number of observations (the larger the better); c) the diversity of variables measured; and d) whether the data set adds to the variety of represented disciplines. Example data sets used in recent offerings of the course include:

**The General Social Survey** (<http://www.norc.uchicago.edu/GSS+Website/>): The General Social Survey (GSS) conducts basic scientific research on the structure of American society. Research questions based on this data set have included: Are individuals with Christian beliefs more likely to report high levels of patriotism? Are there gender differences in attitudes toward censorship?

**Forest Caterpillar Ecology Study:** This study conducted by Michael Singer, Ph.D. (Singer, Mace, & Bernays, 2009) is aimed at understanding food web structure in a forest ecosystem. Research questions based on this data set have included: Do different tree species in the forest host different numbers of caterpillar species? Do different caterpillar species suffer different frequencies of parasitism (i.e. species of wasps or flies that lay eggs in or on a caterpillar)?

**The National Longitudinal Study of Adolescent Health** (Add Health), [http://www.nichd.nih.gov/health/topics/add\\_health\\_study.cfm](http://www.nichd.nih.gov/health/topics/add_health_study.cfm) is a nationally representative study that explores the causes of health-related behaviors of adolescents in grades 7 through 12 and their outcomes in young adulthood. The public access version of the baseline assessment was made available to students in the course. Research questions have included: Do adolescents whose parents are divorced experience higher levels of depression than those from intact families? Is religious affiliation associated with contraception use?

### *Stress conceptual understanding rather than mere knowledge of procedures*

Rather than focusing on rules associated with traditional lists of statistical tools (e.g.,  $z$ -test, one sample  $t$ -test, two sample  $t$ -test, paired  $t$ -test, etc.), we have organized the course according to the decisions and skills involved in statistical inquiry. Basic themes such as measurement and descriptive and graphical representation are covered, as well as more specific inferential methods needed to test hypotheses and/or explore the empirical structure of data. All, however, are introduced as the student's scientific questions dictate their presentation. In this way, students are provided with opportunities to learn to evaluate what tools would be most appropriate for their research question(s) and to engage in decision making. While not all students utilize the entire menu of tools offered in this course, through lecture, on-line materials, and collaboration with peers, they are exposed to a wide variety of methods, and learn to choose and use them flexibly as they are needed. This approach is aimed at building student confidence and their ability to evaluate data and seek out appropriate methods for the questions at hand.

### *Foster active learning in the classroom*

The examples presented to the class are based on individual student experiences, and much of the more detailed instruction takes place "after the fact", that is in the context of the questions and needs that the students' work generates (Kester, Kirschner, & Van Merriënboer, 2004). In this way, like professional scientists, students "decompose their topic, identify key components; abstract and formulate different strategies for addressing it; connect the original question to the statistical framework; choose and apply methods;

reconcile the limitations of the solution; and communicate findings” (Nolan & Temple Lang, 2009). Based on the student’s choice of data, each generates testable hypotheses, conducts a literature review on their topic of interest, works to refine or broaden their research questions based on information they collect, prepares data for analysis (i.e. data management), selects and conducts descriptive and inferential statistical analyses; and evaluates, interprets and presents research findings. These activities are not presented or experienced as distinct stages but rather, as a series of ongoing, interactive tasks.

### *Use technology for developing conceptual understanding and analyzing data*

We believe that the ability to use statistical software packages is a central skill that greatly expands a student’s capacity not only for statistical application, but for engaging in deeper levels of quantitative reasoning. Thus, an important aspect of the course is exposing students to popular statistical software tools that can be used to manage and analyze data. Not surprisingly, opinions differ widely both across substantive disciplines and even within individual departments about the specific statistical software packages that should be taught. Many introductory statistics courses now cover the practical aspects of using a single statistical software package. As noted by (Nolan & Temple Lang, 2009), this exposure is most often targeted at a basic knowledge about the particular package, rather than being used as a platform for the more important goal of conceptual reasoning with data. Our approach is a more general and translatable one that is meant to provide students with flexible skills that transcend the specific software to explore data and formulate and test scientific questions. Specifically, we expose students to the use and translation of four commonly used, broad statistical software packages (SAS, Stata, SPSS and R), focusing on the commonality and patterns that will provide them with a powerful, general viewpoint and more flexible understanding of data management and statistical analysis (Nolan & Temple Lang, 2009). Our choice to introduce students to different software is motivated by the goal of developing transferable skills which in turn allow students to use the “best tool for the job” instead of “getting stuck” with particular software. Importantly, although some common statistical software packages have developed a point and click interface that allows students to bypass more formal logic syntax, we employ the logic syntax-based approaches to statistical computing in an effort to allow students to fully engage in the decision making process of scientific inquiry (e.g. the connection to the logic of data management and the choices made in statistical analysis and general model building). To accomplish this, we have developed translational resources that provide students with appropriate syntax for achieving a host of data management and analytic tasks of use in the pursuit of answers to question of the greatest interest to them.

### *Use assessments to improve and evaluate student learning*

To evaluate students learning, we use several different assessment methods:

**Laboratory assignments:** Students complete 12 lab assignments during the semester with each assignment allowing the instructor to gauge progress in their project.

**The Open Learning Initiative (OLI)** <https://oli.web.cmu.edu/openlearning/> provides pre-existing on-line course packages for introductory statistics that include on-line text and interactive activities. In addition, students are given the opportunity to take non-graded self-assessments to gauge their understanding of the course material and instructors are provided with the same timely feedback on student progress and the level of on-line support that they needed.

**Research Plan:** Students prepare and submit a written research plan that includes a literature review on their research topic, a description of the study methods and an evaluation of the importance of the research question. This is meant to give students and opportunity to receive concrete feedback on their formal scientific reasoning and writing.

**Exams:** Four quarterly in-class multiple choice exams are given in which students are asked to integrate material from lecture, lessons and laboratory experiences.

**Poster/Oral Presentation:** Lab assignments build to the completion of the individual project which is presented and evaluated at the end of the semester as a research poster and oral presentation.

## Course Enrollment and Evaluation

This course was first offered during the fall semester of 2009, and a total of 75 undergraduates were enrolled. A second offering of the course during fall 2010 enrolled 57 undergraduates and a third, during fall 2011 enrolled 98 undergraduates. Each offering of the course enrolled students with a variety of disciplinary interests and backgrounds. Among those who had declared a major (60.0% of students), 15% were majoring in a natural science or mathematics, 64% in a social science and 21% in the humanities. Satisfaction with the course was high and did not differ significantly among students pursuing majors in each of these areas.

Students completed surveys at the beginning and end of the semester (i.e. pre and post).. The pretest was completed prior to the end of the first week and the posttest during the last week of the semester. Each survey took approximately 10-15 minutes to complete. Students were informed that their participation in this study was confidential and voluntary with no impact on their course grade.

Based on the posttest, 72.4% of students rated their work in the course as rewarding or very rewarding and 83.6% felt that their effort in completing the semester long project was worth the skills that they developed. Further, more than half of the students (55.2%) rated the course as more useful than other college courses they had taken (40% felt that it was similarly useful and fewer than 5% felt that it was not as useful). At the end of the course nearly three-quarters of student believed that they were likely or very likely to use the research skills that they had been exposed to again in the future. More than 90% of students felt that they would recommend the course to others (55.5% said they would definitely recommend the course and 36.5% said they would probably recommend it).

Projects were judged to be rewarding or very rewarding by 57.6% of students and 6.8% felt it was the most rewarding project they had ever completed for a course. Less than 2% of students felt that their project was not rewarding. More than 60% of students felt that they were likely or very likely to use the research skills that they had learned in the course again; 9% felt they would definitely use them again and only 6% felt that it was not likely they would be used again. Nearly half of the students (47%) agreed that they would take another statistics course in the future and only 9.1% of students felt that they would not recommend this course to others. In summary, the course appears to be succeeding in fostering exposure, understanding and appreciation of the relevance of statistical inquiry.

A comparison of student characteristics in our project-based course vs. the traditional introductory statistics course offered through our Math department is presented in Table 2. While both courses show similarly high rates of female enrollment (>60%), our project-based course attracted significantly higher rates of ethnically under-represented students (i.e. Black and Hispanic students) compared to the math course.

**Table 2. Student Characteristics for Project-Based and Traditional Introductory Statistics Sections.**<sup>1</sup>

	<b>Project-Based Statistics</b> <i>n</i> = 206	<b>Math Statistics</b> <i>n</i> = 188	<b>Statistics</b>
Ethnicity			
Hispanic	25 (12.1%)	8 (4.3%)	$\chi^2=7.9$ , 1 df, $p<.005$
Black	27 (13.1%)	14 (7.5%)	$\chi^2=3.4$ , 1 df, $p<.06$
Asian	38 (18.5%)	34 (18.1%)	n.s.
White	112 (54.4%)	123 (65.4%)	$\chi^2=5.0$ , 1 df, $p<.03$
Under-represented (Black/Hispanic)	51 (24.8%)	22 (11.7%)	$\chi^2=11.1$ , 1 df, $p<.0009$
Gender (% Female)	130 (63.1%)	122 (64.9%)	n.s.

<sup>1</sup>Data through fall semester 2011; n.s. = non-significant

Of the numerous reasons indicated for early departure from science and engineering programs, the most often cited reason is unsuitable or uninspiring pedagogical practices (Handelsman, 2005; Seymour, Hunter, Laursen, & Deantoni, 2004). Importantly, the response to our project-based course by underrepresented students was excellent. Based on the anonymous semester end survey, 72.0% of Black students and 81.2% of Hispanic students were interested in taking a follow-up course and a large proportion of students rated the course as more useful than others they had taken in college (Black, 45.8%, Hispanic, 52.4%, Asian, 51.4% and White 58.1%). Further, many felt that they were very likely to use the skills they have learned in the future (Black, 44.0%, Hispanic, 42.9%, Asian, 46.0% and White 44.4%). A total of, 70.8% of Black students and 71.4% of His-



panic students indicated that they would definitely recommend the course to their peers (compared to 49.1% of White students and 62.2% of Asian students).

Evaluations of the semester-long research project were equally positive. When asked if the effort involved in completing the research project was worth the skills developed, 79.2% of Black students and 86.4% of Hispanic students judged that their skills were worth or very worth the effort involved, compared to 86.5% of Asian students and 83.0% of White students. Sixty-eight percent of Black students and 77.3% of Hispanic students found the project rewarding or very rewarding, compared to statistically similar rates among Asian (77.8%) and White (69.5%) students.

## Summary

The creation of a multidisciplinary project-based curriculum and the integration of maximally supportive resources takes advantage of students' natural curiosity and provides a common language for approaching questions across numerous disciplines. Importantly, the materials and supportive training infrastructure for this course have been designed in such a way that other instructors can easily follow the framework and utilize the newly developed resources. Our goal is not to deliver a course whose success is tied to individual instructors being facile in all statistical software or methods. Instead, we are developing a strong infrastructure of resources and expertise that can be drawn on by both instructors and students. In this way, the course not only develops vital student expertise, but also develops instructor expertise in terms of their ability to mentor students, engage in the use of statistical and methodologic vocabulary across disciplines and develop expertise across leading statistical software packages.

This project-based course provokes students to encounter (and struggle with) the central concepts and principles not only within the discipline of statistics, but also with the discipline that their chosen research reflects. Although our model focuses on statistics education, the emphasis on authentic real-world activities with the goal of sparking interest and enthusiasm (BIE, 2012) can be achieved in curricular content as diverse as science (Kubiak & Vaculov, 2011) and foreign language (Danan, 2010) instruction. In addition to the specific skills most directly emphasized, project-based courses provide students with experience in communication, organization and time management (BIE, 2012). We believe that our course can benefit other universities not only through dissemination of our model and experiences, but by making our newly developed resources widely available. We are happy to share our course materials with others and encourage faculty to consider integrating project-based course content. <http://www.wesleyan.edu/qac/curriculum/>

## Acknowledgements

This research was supported by grant 0942246 from the National Science Foundation, Transforming Undergraduate Education in Science, Technology, Engineering and Mathematics (TUES) and the Lauren B. Dachs Grant in Support of Interdisciplinary Research in the Social Impacts of Science. We appreciate the generous efforts of our colleagues and advisors who contributed to the development of this course: Drs. George Cobb, Lisa Har-

low, Daniel Long, Michael Singer, Wendy Rayack, Erika Fowler, John Kirn, and Marc Eisner. We also thank Mr. Michael Whitcomb for his assistance with portions of the evaluative data, Dr. Chien-Ti Lee for her work managing data for the course and Dr. Ma-yumi Gianoli for her contributions as a laboratory instructor.

## References

- Buck Institute for Education (BIE) and Boise State University, Department of Educational Technology; <http://pbl-online.org/> accessed April 1, 2012.
- Danan M. (2010). Dubbing projects for the language learner: a framework for integrating audiovisual translation into task-based instruction. *Computer Assisted Language Learning*, 23 (5) 441-456.
- Franklin, C., & Garfield, J. (2006). The Guidelines for Assessment and Instruction in Statistics Education (GAISE) Project: Developing Statistics Education Guidelines for Pre K-12 and College Courses., In G. F. Burrill (Ed.), *Thinking and Reasoning about Data and Chance: Sixty-eighth NCTM Yearbook* (pp. 345-375 ). Reston, VA: National Council of Teachers of Mathematics.
- Handelsman, J. (2005). More women in science. *Science*, 309, 1190-1191.
- Harada, V. H., & Yoshina, J. M. (2004). Moving from rote to inquiry: Creating learning that counts. *Library Media Connection*, 23, 22–24.
- Hickey, D. T., Kindfield, A. C. H., Horwitz, P., & Christie, M. A. (1999). Advancing educational theory by enhancing practice in a technology-supported genetics learning environment. *Journal of Education*, 181, 25–55.
- Hickey, D. T., Wolfe, E. W., & Kindfield, A. C. H. (2000). Assessing learning in a technology-supported genetics environment: Evidential and consequential validity issues. *Educational Assessment*, 6, 155–196.
- Kester, L., Kirschner, P., & Van Merriënboer, J. (2004). Timing of Information Presentation in Learning Statistics. *Instructional Science*, 32(233-252).
- Kubiatko, M., & Vaculova, I. (2011) Project-based learning: characteristic and the experiences with application in the science subjects . *Energy Education Science and Technoogy – Social and Educational Studies* 3(1-2), 65-74.
- Langer, J. A. (2001). Beating the odds: Teaching middle and high school students to read and write well. *American Educational Research Journal*, 38, 837–880.
- Lynch, S., Kuiper, J., Pyke, C., & Szesze, M. (2005). Examining the effects of a highly rated science curriculum unit on diverse students: Results from a planning grant. *Journal of Research in Science Teaching*, 42, 921–946.
- National Academies. (2004). *Facilitating Interdisciplinary Research*. Washington D.C.: The National Academies Press.
- Nolan, D., & Temple Lang, D. (2009). Approaches to Broadening the Statistics Curricula. In M. C. Shelley, L. D. Yore & B. Hand (Eds.), *Quality Research in Literacy and Science Education*. London, UK: Springer.
- Seymour, E., Hunter, A., Laursen, S., & Deantoni, T. (2004). Establishing the benefits of research experiences for undergraduates in the sciences: First findings from a three-year study. *Science Education*, 88, 493-534.

Singer, M. S., Mace, K. C., & Bernays, E. A. (2009). Self-medication as adaptive plasticity: Increased ingestion of plant toxins by parasitized caterpillars. *Public Library of Science*, 4(e4796).

Wuchty, S., Jones, B., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827), 1036-1039.