THE IMPACT OF USING REAL-WORLD DATA ANALYSIS IN APPLIED BUSINESS STATISTICS COURSES

Rebecca J. Davis, Stephen F. Austin State University
Beverly Mendoza, Stephen F. Austin State University

ABSTRACT

An increasing number of employers are seeking college graduates with data analysis skills. However, research suggests that students have an aversion to statistics courses and avoid utilizing statistical tools in the workplace. Project-Based Learning has been proven to increase student engagement and achievement, particularly in courses with a heavy mathematical component. To bridge the gap between employer wants and student sentiment related to statistics, instructors of applied business statistics at a regional comprehensive public university implemented a group project that utilized real-word data to answer a research question. This paper examines if and through what channels this group project impacted student performance and knowledge using multiple linear regression models. Results indicate that inclusion of the group project is associated with higher final grades but lower average exam scores, while controlling for student ability using student cumulative grade point average as a proxy. Focusing on the students that completed the project, a higher project grade corresponds to both improved final grades and average exam scores, implying that the project solidifies learning of course concepts.

Keywords***:*** *Project-Based Learning, applied business statistics, business statistics, Microsoft Excel, higher education*

INTRODUCTION

Data analysis has recently become a desired skill requirement for job seekers. As part of preparing undergraduate students for the work force, the applied business statistics course at a regional comprehensive public university has implemented a group project that enables students to apply their skills and knowledge in answering real-world questions. The project simulates real world data analysis using publicly available datasets and applying statistical concepts. In our paper, we analyze how the project impacts college students' statistical learning.

Research has shown that students express negative feelings towards statistical courses (Deckard, 2017). Some even find statistical analyses to be unusable in evaluating information and tend to avoid it in their professional lives (Petocz & Reid, 2005). To combat the negative notion against statistics, a hands-on approach is essential in developing statistical thinking through practical applications. In line with Bailey, Spence, & Sinn (2013) and Dierker et al.

(2018), the assigned group project allows students to discover the practical uses of data analysis and experience it using real-world data.

The applied statistics course implements the Project-Based Learning (PjBL) approach by assigning a group project that is completed throughout the semester. PjBL has risen in popularity in recent decades since it closes the gap between the understanding of concepts and their application. It results in higher levels of student engagement (Wurdinger, Haar, Hugg, & Bezon, 2007) and larger positive effects on student academic achievement (Chen & Yang, 2019). Comparing it with traditional instruction, PjBL is not more demanding in terms of resources and time, and it can even be implemented with few resources available inside the classroom (Al-Balushi & Al-Aamri, 2014). The main principle behind it is to produce a tangible product using real world problems over a period of time (Oguz-Unver & Arabacioglu, 2014).

The application of PjBL in statistics courses creates a positive environment for students to practice statistical techniques using real-world data (Dierker, et al., 2018). Through the project, students can foster deep-level learning and important skills for professional practice (Fernandes, 2014) and connect classroom concepts with the real world (Efstratia, 2014).

Each group is assigned a research question along with a corresponding dataset. The project is comprised of two main steps: Data visualization and regression analysis. The first step is submitted in the middle of the semester in order for students to receive feedback regarding their understanding of the assigned dataset; while the second step is due at the end of the semester which culminates their understanding of statistical concepts.

Studies on the effect of PjBL in higher education are currently limited (Guo, Saab, Post, & Admiraal, 2020). Lee, Blackwell, Drake, & Moran (2014) point out that most of the research focuses mainly on K-12 education, while Kokotsaki, Menzies, & Wiggins (2016) uncover that most studies of PjBL in higher education are centered around engineering courses. Our paper extends the literature by examining the impact of PjBL in a college environment, specifically for a business course.

Through a series of regression analyses, we find that adding a semester-long project increases final course grades but detracts from midterm exam scores. Particularly, we observe that the project is associated with an average increase of 3.11 percentage points in a student's final grade. However, students who were required to complete the project experienced a decrease in midterm grades by 2.57 percentage points on average, compared to students who did not have to take the project. An explanation could be that time spent on the project is substituted away from studying for the exams.

Focusing only on students who have completed the project, we reveal further evidence of the positive effect of the project on learning. Examining the effect of the average project grade, we find that an additional percentage point increase in the project grade is associated with an average increase of 0.29 percentage points in their final course grade and 0.14 percentage points on their average exam grade. Throughout our analyses, we account for higher-achieving students by controlling for their cumulative GPA and credit hours in our regression models.

Our paper proceeds as follows. The following section details the institutional background surrounding the project and class structure. The section after that describes the data used with another section explaining our empirical methodology and results afterwards.

BACKGROUND

Applied statistics is a required undergraduate course for all business-related majors (e.g., economics, finance, management, marketing, etc.). It covers topics ranging from defining and collecting data to analyzing it, utilizing techniques such as multiple linear regression models. Students typically take this course in their junior or senior year, though some sophomores take the course.

Classes generally met in person twice a week (150 instructional minutes per week), except during the COVID-19 pandemic. In Spring 2020, classes were conducted face-to-face until the middle of the semester when both the federal and state governments recommended individuals to limit social gatherings to ten people or less (The White House, 2020). As a response, university administration dictated that students were not to return to campus after their spring break; instead, classes resumed remotely through online live-streaming or recorded lectures. The following academic year (2020-2021) was taught in a hybrid environment[1] following the pandemic restrictions (The White House, 2020). Beginning Fall 2021, instruction was resumed normally, and restrictions were alleviated. Our analysis in the subsequent sections accounts for the changes resulting from the pandemic.

Historically, students were only required to complete homework assignments and exams. As of Fall 2021, a semester-long group project has been implemented in the applied statistics course. Students taking the course in Fall 2021 and each consecutive semester after have been required to complete the group project.

Project

The semester-long group project enables students to perform data analysis techniques, ranging from data visualization to regression analysis. Its goal is to simulate real-world data analysis by applying statistical concepts to publicly available data using Microsoft Excel.[2] Through addressing their assigned research question, students were tasked to clean, organize, visualize, and analyze their dataset. Skills learned from the course include, but are not limited to, the use of Excel pivot tables, charts, formulas, and regression analysis.

Table 1
Examples of Research Questions and their Corresponding Datasets

RESEARCH QUESTION	CORRESPONDING DATASET
What determines the market value of a home?	American Housing Survey
Do crime and crime prevention measures impact student grades?	National Crime Victimization Survey
What factors influence grades for K-12 students?	National Household Education Survey

[1] Half of the students attended in person, while the other half attended remotely via Zoom.

[2] Most middle-skilled jobs require at least basic understanding of Microsoft Excel (Formby, Medlin, & Ellington, 2017). For conciseness, the rest of the paper will refer to Microsoft Excel as Excel.

Students are allowed to choose their own groups and research question/dataset that they were interested in. Example research questions and sources of data are provided in Table 1. Permitting them to self-select into groups increases student engagement and motivation towards the project (Crossouard, 2012; Wurdinger, Haar, Hugg, & Bezon, 2007). Students who did not choose their groups by the deadline are randomly assigned into one. To uphold accountability and responsibility within a group, students are required to complete peer evaluations of their group members.

The project has two main steps: 1. Data visualization and 2. Regression analysis. Each step is submitted separately but is comprehensive in nature; particularly, the regression analysis part can be successfully completed after having a good understanding of their data from the data visualization part. Grading each part separately provides students with ample feedback to learn from their application of statistical techniques and improve upon them. The entire project, including Step 1, Step 2, and peer evaluations accounted for 20 percent of a student's overall course grade.

The first step requires students to clean, organize, and visualize their data. They are asked to determine each variable by its type (numerical or categorical), measure (continuous or discrete), and scale (interval, ratio, nominal, or ordinal). Students are also tasked to provide intuition on how each variable relates to the research question. Lastly, students calculate summary statistics and describe the distribution for each of the variables included in their dataset.

In the second step of the project, they are tasked to analyze data using multiple linear regression models. They are expected to interpret each regression coefficient and perform statistical significance tests (T-tests and F-tests). After estimating several regression models, students determine which one is their preferred model for answering their assigned research question. Students, then, explain their process in choosing their preferred model, including why they believe it is the model that best answers the research question. In justifying their model, they would need to consider the adjusted coefficient of determination (R^2) and statistical significance of each variable. Finally, students comment on the sample size, irrelevant, and omitted variables.

Because of the free-rider[3] problem, students are required to submit peer evaluations of their group members, which is factored into their project grade (Levin, 2003). Peer evaluations were initially submitted at the end of the semester (after students finished Step 2 of the project); however, this enabled less accountability from some students during Step 1 of the project. Therefore, in subsequent semesters, students submit peer evaluations at both stages of the project. To further minimize the free-rider problem, the course imposed the rule that a student's project grade is subject to zero if each of the group members unanimously graded the student

[3] Free-riders include one or more members of a group limiting the work that they contribute, knowing they will benefit from the work of others in the group regardless of how much work they provide.

zero across all categories of the peer evaluation.[4] Although the rule is strict, it has rarely been applied.

Homework

In addition to the project, students complete homework assignments and exams through Pearson's MyLab Statistics,[5] which is a required purchase for all students. The purchase requirement is consistent across all classes of applied business statistics, using the same textbook (Levine, Stephan, & Szabat, 2021) and corresponding online learning platform (MyLab Statistics).

Homework assignments are generally consistent across all classes of applied business statistics. Depending on the instructor and semester, students are assigned between seven to ten homework assignments, with up to two of the lowest homework scores dropped and omitted from the calculation of their overall grade.

In our empirical investigation, detailed in the subsequent section, we closely examine four homework assignments (Chapters 2, 3, 13, and 14) as they are most closely related to the project. Homework assignments on Chapters 2 and 3 are associated with Step 1 of the project, while assignments on Chapters 13 and 14 are related to Step 2 of the project. Chapter 2 covers Excel work, including building pivot tables and charts. Chapter 3 reviews basic statistical calculations, such as mean, median, mode, variance, standard. Chapter 13 is concerned with simple linear regression, while Chapter 14 is comprised of multiple linear regression.

Exams

Students are required to complete three exams and one cumulative final exam. Exams are administered using the MyLab Statistics online platform throughout the entire study period. Before the introduction of the project, students took the first three regular exams outside of class using MyLab Statistics and took the cumulative final exam in class as the only proctored exam. Note, even the cumulative final exam was administered using MyLab Statistics but the physical location the student took the exam changed to in class (a computer lab) with the instructor present. This was true for both instructors.[6] However, during the same semester the project was implemented, the cumulative final exam was made optional for students to replace their lowest

[4] The peer evaluation form lets students grade each of their group members based on five categories: contribution, communication, responsibility, cooperation, and participation. A group member's grade would be the average of ratings across the five categories.

[5] Pearson MyLab Statistics is an online learning platform that includes textbook author content and assignable algorithmic exercises for extra practice (Pearson, 2023).

[6] For the Spring 2020 semester, where half the course was taught in person and the other half online, the cumulative final exam was proctored using Zoom. Students had to Zoom with the instructor while they took the cumulative final exam from home. The same was true for the hybrid academic year, Fall 2020 – Spring 2021.

exam score.[7] Additionally, one instructor began proctoring all exams, meaning students took the tests utilizing MyLab Statistics in class with the instructor present. The other instructor kept a similar format to before the project was implemented and had students take the first two exams outside of class using MyLab Statistics and proctored the third exam in the classroom, still using MyLab Statistics as the online platform.

DATA

Our dataset primarily consists of grades and student performance on assignments that were collected as part of electronic grade books for each section of applied statistics. It contains grades from 1,202 students across 32 course sections taught by two instructors, spanning over a five-year period (Fall 2018 – Fall 2022).[8] Out of the 1,202 students, 774 were not subject to the project, meaning the remaining 428 students were required to complete the project as a part of the course. Summary statistics for student grades and demographic controls are displayed in Table 2. After receiving IRB exemption, demographic data was provided by the university and matched to the corresponding student via student identification numbers. Once merged, the data were de-identified for analysis.

A typical student in our sample is 21 years old and attempts an average of 14.04 hours during the semester they complete the applied statistics course. There are more males (60.00 percent) than females (40.00 percent) in the analytic sample. The breakdown of race shows 60.00 percent of students self-report as white, while another 20.00 percent self-report as Hispanic and 12.00 percent self-report as black or African American. The average cumulative grade point average (GPA) was 2.88 during the semester the student took the course.

To control for differences in instructor experience, ability, and management of course assignments, we include a dummy variable equal to one for Instructor 1 and equal to zero otherwise (for Instructor 2). Instructor 1 (2) was responsible for teaching 63.00 (37.00) percent of the sample. The time variable is equal to one for the first semester included in the sample (Fall 2018) and increases by one for each consecutive semester, meaning if a student took the course Spring 2019, the time variable is equal to two, etc. The time variable captures differences in the course from semester to semester, but it also allows us to control for semesters that the COVID-19 pandemic altered the modality of the course.

The average final grade for students in the sample is 78.08 percent, which is a C letter grade. Students scored higher on average for Chapters 2 and 3 homework than on Chapters 13 and 14 homework, which corresponds with the difficulty of the concepts included in the respective chapters. Chapter 3 homework saw the highest average grade of 78.82 percent, while students had the lowest average grade of 62.54 percent on Chapter 14 homework. A similar

[7] The main reason for this change is that the course content was fully covered by the time of the third exam, meaning there was no new information tested on the cumulative final exam.

[8] The sample excludes students that dropped or withdrew from the course.

pattern arises for exams. Students earned an average of 78.08 percent on Exam 1, 77.38 percent on Exam 2, and 72.12 on Exam 3 to make their average exam grade 75.85 percent.

Table 2
Summary Statistics

	Mean	Standard Deviation	Minimum	Maximum	Observations
Panel A					
Final Grade	78.08	15.61	2.33	103.00	1202
Chapter 2 HW	73.49	29.34	0.00	100.00	1202
Chapter 3 HW	78.82	29.66	0.00	100.00	1202
Chapter 13 HW	72.01	31.78	0.00	100.00	1202
Chapter 14 HW	62.54	32.23	0.00	100.00	1133
Exam 1	78.06	15.61	0.00	101.04	1202
Exam 2	77.38	19.66	0.00	111.38	1202
Exam 3	72.12	21.06	0.00	100.00	1202
Average Exam Grade	75.85	15.52	0.00	101.88	1202
Average Project Grade	81.72	11.83	28.70	100.00	428
Average Peer Evaluation Score	87.05	24.54	0.00	102.00	428
Panel B					
American Indian or Alaskan Native	0.01	0.08	0.00	1.00	1202
Asian	0.01	0.12	0.00	1.00	1202
Black or African American	0.12	0.33	0.00	1.00	1202
Hispanic	0.20	0.40	0.00	1.00	1202
International	0.02	0.13	0.00	1.00	1202
Two or More	0.04	0.20	0.00	1.00	1202
Unknown or Not Reported	0.00	0.07	0.00	1.00	1202
White	0.60	0.49	0.00	1.00	1202
Gender (Female=1)	0.40	0.49	0.00	1.00	1202
Instructor (Instructor 1=1)	0.63	0.48	0.00	1.00	1202
Time	5.28	2.40	1.00	9.00	1202
Age	21.34	3.05	18.00	57.00	1202
Cumulative GPA	2.88	0.63	0.00[9]	4.00	1202
Attempted Hours	12.04	2.27	3.00	21.00	1202

Note: Panel A lists variables associated with student grades. Panel B displays variables related to student controls.

[9] There are ten observations with a cumulative GPA of zero. Eight of them are transfer students (listed as a sophomore, junior, or senior), one is an international student attending for a semester (listed as a freshman), and one is a first-time student taking classes necessary for the master's in professional accountancy program. Keeping these observations in the analysis potentially alters the predictive power of the cumulative GPA variable, but after estimating the models included in the remainder of this paper with and without these ten observations, the main results hold. Thus, we have kept them in the analysis. As a point of reference, the lowest cumulative GPA, ignoring the ten observations with a zero cumulative GPA, is 1.18.

For the students that were subject to the project, the average Step 1 and Step 2 project grade was 81.72 percent. The average peer evaluation score was even higher at 87.05 percent, indicating that students seemed to work well in a group setting according to their peers. Notice that some grade categories have maximum scores higher than 100 percent. That is due to the inclusion of extra credit opportunities in all semesters for both instructors. Specifically, one student is responsible for the maximum final grade and average exam grade reported in Table 2. Three students earned a grade higher than 100 percent on Exam 1, while another three students earned a grade higher than 100 percent on Exam 2 due to extra credit opportunities. The maximum peer evaluation score was 100 for all students, except one in which that student's groupmates insisted this student should earn additional points and the instructor obliged. Thus, the maximum average peer evaluation score that is higher than 100 is associated with one student.

EMPIRICAL METHODS & RESULTS

Employing the data described in the previous section, we estimate several linear regression models to determine if the inclusion of a group project enhances student learning and improves final grades. Since the group project occurs in two steps, we further explore which step of the project has the larger impact on student success. Importantly, the models and results presented in the remainder of this section highlight correlations of the project with learning outcomes. Because all students were subject to the project after its introduction, there is no valid control group that was not subject to the project after it was implemented in the course. Therefore, there is no empirical strategy to exploit that would provide causal impacts. The control group described previously refers to those students that completed the course before the project was implemented. Thus, the results below are correlations.

Project Impact on Grades

The impact of a group project on grades was determined by estimating the following regression model.

$$Y_i = \alpha + \beta Project_i + C_i + \epsilon_i \qquad (1)$$

The dependent variable in Equation 1 is either the final grade or the average exam grade for student *i*. $Project_i$ is a binary indicator equal to one if the student was subject to the project and zero otherwise. C_i contains controls for race, gender, instructor, time, age, cumulative GPA, and attempted hours. The coefficient of interest, β, describes the average effect of the project on final grades or average exam scores. Table 3 displays the results from estimating Equation 1.

A student that was subject to the project is associated with an average increase in their final grade of 3.11 percentage points compared to a student that was not responsible for the project. This result is statistically significant at the 5 percent level. If a student self-reported as

American Indian, Alaskan Native, or International correlated to significantly higher final grades on average compared to students that self-reported as white. Students that enrolled in a section taught by Instructor 1 saw higher grades on average. The coefficient on time indicates that for each additional semester taught, average final grades decreased by 0.61 percentage points on average. We believe this result is due to the negative impact of COVID-19 on student motivation (Borgaonkar, Sodhi, Vijayabalan, & Nair, 2021). Unsurprisingly, increasing a student's cumulative GPA by one entire point (e.g., going from a 2.0 cumulative GPA to a 3.0) increases the average final grade by 12.43 percentage points. Cumulative GPA proxies for student ability. Thus, a higher cumulative GPA can be indicative of a student's academic ability as well as their commitment to learning and dedication to coursework.

Table 3
Estimates of Project Impact on Grades – Indicator

	Final Grade	Average Exam Grade
Project	3.1060** (1.41)	-2.5660* (1.43)
American Indian or Alaskan Native	9.9305** (5.01)	6.6368 (5.08)
Asian	9.1351*** (3.23)	8.6596*** (3.28)
Black or African American	-1.5563 (1.21)	-1.9797 (1.22)
Hispanic	1.2598 (0.99)	0.6853 (1.01)
International	5.0698* (3.00)	5.6164* (3.04)
Two or More	-0.7047 (1.96)	-1.2686 (1.99)
Unknown or Not Reported	-1.0779 (5.43)	-4.7935 (5.51)
Gender (Female=1)	-0.2182 (0.79)	-1.2977 (0.80)
Instructor (Instructor 1=1)	3.9113*** (0.83)	3.2937*** (0.84)
Time	-0.6101** (0.29)	0.1615 (0.29)
Age	-0.0761 (0.13)	-0.0899 (0.13)
Cumulative GPA	12.4349*** (0.63)	11.9241*** (0.63)
Attempted Hours	0.0038 (0.17)	0.0556 (0.18)
R^2	0.299	0.271
Observations	1202	1202

Note: The dependent variable in the first column is final grade. The dependent variable in the second column is the average score on Exams 1-3. The variable of interest is an indicator equal to 1 if the student was subject to the project. Standard errors are reported in parentheses below the coefficients. * p<.10, ** p<.05, *** p<.01.

Alternatively, average exam scores dropped once the project was introduced to the course. Specifically, if a student was required to complete the project, their average exam score decreased by 2.57 points on average. Importantly, Step 1 of the project was due in between Exam 1 and Exam 2, while Step 2 was due after Exam 3 but within the same week. This was true for all semesters and instructors. It could be that students subject to the project were shifting time away from studying for exams towards project completion. Student controls have similar impacts on average exam scores as they did on final grades.[10]

Based on our findings, the project increases final grades but detracts from exam scores. The project is a group effort, and it may be the case that the project acts as a grade cushion, allowing students to perform worse on exams as a higher group project counteracts the exam grade category. This is not ideal, but if the project improves overall student learning, there is an argument that the inclusion of the project was beneficial. We explore this possibility in the next section.

Project Impact on Learning

If the group project increases learning, a student's average project grade should increase both their overall grade as well as their average exam grade. This was determined by estimating the following regression model. Note, the model is conditional on the student being subject to the project.

$$Y_i = \alpha + \beta AverageProjectGrade_i + C_i + \epsilon_i \quad (2)$$

The dependent variable in Equation 2 is either the final grade or the average exam grade for student i. $AverageProjectGrade_i$ is calculated as the average project grade across Steps 1 and 2. C_i contains controls for race, gender, instructor, time, age, cumulative GPA, and attempted hours. Additionally, C_i includes a control for average peer evaluation score to account for individual student effort in a group setting. The coefficient of interest, β, describes the average effect of the project on final grades or average exam scores only for students that were subject to the project. Table 4 shows the results from estimating Equation 2.

Results indicate that each additional one percentage point increase in a student's average project grade is associated with an average increase of 0.29 percentage points in their final grade and 0.14 percentage points in their average exam grade. The impact of average project grades on final grades is expected since the project is a part of the calculation for a student's final grade. However, the fact that higher average project grades also enhance average exam scores could imply that application of concepts covered during lecture via the project solidify learning and ultimately increase average student performance on exams. Conversely, the positive coefficient associated with average project grade could be picking up that higher achieving students have

[10] Self-identified Asian and international students perform better than self-identified white students. Instructor 1 students had higher average exam grades on average, and students with higher cumulative GPAs earned higher average exam grades.

higher average project scores as well as higher average exam grades. We do not believe that is the case, because we control for cumulative GPA as a measure of student achievement.[11] For this reason, we believe these results imply the project enriches student learning and that spreads to exam performance.

In Equation 2, the project is measured by calculating the average project grade across Step 1 and Step 2. We are interested in understanding how each step of the project influences individual exams. Each step of the project is a potential pathway for applying course concepts to increase learning reflected in grades. Though, only Step 1 of the project has subsequent exams that occur after the project deadline. The next section explores Step 1 of the project.

[11] Again, student controls have similar effects on final grades and average exam scores as previous model estimations. Even though the project is a group effort, students that garner higher average peer evaluation scores increase both final grades and average exam grades, signifying the importance of being a team player.

Table 4
Estimates of Project Impact on Grades – Average Project Grade

	Final Grade	Average Exam Grade
Average Project Grade	0.2912*** (0.04)	0.1372** (0.05)
Average Peer Evaluation Score	0.2314*** (0.02)	0.2284*** (0.03)
American Indian or Alaskan Native	3.6035 (9.47)	5.0783 (12.50)
Asian	8.5168** (3.63)	10.3425** (4.80)
Black or African American	-1.8669 (1.46)	-3.1304 (1.93)
Hispanic	1.7480 (1.16)	1.5071 (1.53)
International	9.6637** (4.30)	11.9019** (5.67)
Two or More	-1.4972 (2.42)	-2.9461 (3.20)
Unknown or Not Reported	6.1652 (6.70)	8.9749 (8.85)
Gender (Female=1)	0.0192 (0.96)	-0.3436 (1.27)
Instructor (Instructor 1=1)	3.6446*** (0.92)	3.8343*** (1.22)
Time	0.1511 (0.57)	0.6018 (0.75)
Age	-0.1974 (0.14)	-0.2571 (0.18)
Cumulative GPA	7.3503*** (0.72)	8.9932*** (0.95)
Attempted Hours	0.0487 (0.21)	0.1707 (0.27)
R^2	0.562	0.428
Observations	428	428

Note: The dependent variable in the first column is final grade. The dependent variable in the second column is the average score on Exams 1-3. The variable of interest is the average project grade across two steps. Estimates are conditional on the student being subject to the project. Standard errors are reported in parentheses below the coefficients. * p<.10, ** p<.05, *** p<.01.

Step 1 Project Impact on Grades

The impact of Step 1 of the group project on grades was determined by estimating the following regression model. Note, this is conditional on the student being subject to the project.

$$Y_i = \alpha + \beta Step1ProjectGrade_i + C_i + \epsilon_i \quad (3)$$

The dependent variable in Equation 3 is either Exam 2 or Exam 3 grade for student *i*. $Step1ProjectGrade_i$ is the grade earned on Step 1 of the project for student *i*. C_i contains the same controls included in Equation 2. The coefficient of interest, β, describes the average effect of Step 1 of the project on Exam 2 or Exam 3 only for students that were subject to the project. Table 5 illustrates the results from estimating Equation 3.

Table 5
Estimates of Project Impact on Grades – Step 1 Project Grade

	Exam 2	Exam 3
Step 1 Project Grade	0.1238 (0.08)	0.0421 (0.08)
Average Peer Evaluation Score	0.2199*** (0.04)	0.3258*** (0.04)
American Indian or Alaskan Native	-5.8794 (17.46)	12.6036 (17.96)
Asian	12.2485* (6.70)	9.5984 (6.89)
Black or African American	-4.2053 (2.70)	-1.9804 (2.77)
Hispanic	1.7184 (2.14)	1.7625 (2.20)
International	13.4508* (7.92)	13.2528 (8.14)
Two or More	-0.9526 (4.47)	-7.8620* (4.59)
Unknown or Not Reported	13.2556 (12.35)	1.2321 (12.70)
Gender (Female=1)	-2.9606* (1.77)	1.6975 (1.82)
Instructor (Instructor 1=1)	8.3218*** (1.70)	3.9130** (1.75)
Time	1.1530 (1.04)	0.8430 (1.07)
Age	-0.0670 (0.25)	-0.5501** (0.26)
Cumulative GPA	9.2486*** (1.32)	10.5770*** (1.36)
Attempted Hours	0.3723 (0.38)	0.3774 (0.39)
R^2	0.299	0.363
Observations	428	428

Note: The dependent variables are Exam 2 and Exam 3, which occurred after Step 1 of the project was due. The variable of interest is the Step 1 project grade. Estimates are conditional on the student being subject to the project. Standard errors are reported in parentheses below the coefficients. * p<.10, ** p<.05, *** p<.01.

The exams evaluated in this section occur after Step 1 of the project. Results included in Table 5 show that the first step of the project is not correlated with subsequent exams, implying pass-through of learning from the project is associated more with Step 2 than Step 1. This result is intuitive as Step 1 tasks on the project are related to basic definitions, Excel skills, and calculating descriptive statistics. Further, those concepts are a review from prerequisite courses.

Information tested on Exams 2 and 3 are methods of analyzing data like hypothesis testing and regression analysis.

Results in Table 4 show the project increases average exam scores, and suggests students learn the course content not only through lecture and homework but also through application by completing the project. Results in Table 5 imply said learning occurs more so for Step 2 of the project, which is associated with regression analysis, than for Step 1, which is connected to visualizing and organizing data.

Robustness Checks

The COVID-19 pandemic impacted Spring 2020, Fall 2020, and Spring 2021 semesters. In the semester directly after COVID-19 restrictions were removed, the semester-long group project was introduced to the business statistics course. One potential issue with the results presented previously in the paper is that the introduction of the project and the end of COVID-19 restrictions occurred simultaneously. Given the current empirical strategy, we are able to control for differences in student cohorts from semester to semester as well as instructor teaching proficiency using the time variable. A different way to control for these issues that also disentangles each semester's impact on student outcomes is to utilize dummy variables for each semester in the sample, leaving out one semester in the analysis to avoid perfect multicollinearity. The key benefit of employing the dummy variables to control for time effects is that it also controls for the differential impact of each semester, including those that happen before, during, and after the COVID-19 pandemic, on student outcomes.

After re-estimating Equations 1, 2, and 3 with the semester dummies instead of the individual time variable, results are relatively similar to the original specifications.[12] Signs and statistical significance remain the same. Magnitudes vary slightly. Appendix Tables A.1-A.3 display comparisons of original specifications and robustness checks.

CONCLUSION

This paper has shown how students respond to the inclusion of a group project in an applied business statistics course. Students that were responsible for completing the project saw higher grades but lower average exam scores than their counterparts that did not have to carry out said project. One interpretation of this finding is that students taking the course in a semester with the project substituted study time away from exams towards the project. Another interpretation is that students were able to put forth less effort in the course, because the project was completed by a group and free-riding could occur.

[12] Equation 1 includes all 1,202 observations in the analytic sample as well as all semesters, Fall 2018 – Fall 2022. The semester dummy that is left out of the model estimation is Fall 2018. Equations 2 and 3 solely focus on students that were subject to the project, which started in Fall 2021. Therefore, the semester dummy that is left out of the model estimations is Fall 2021.

By focusing only on the students that completed the project, we found that earning a higher project grade was associated with both a higher average exam grade and final course grade. Notably, we controlled for student ability and achievement by including cumulative GPA in the analysis. While our results showing higher average project grades increase final grades are unsurprising, the fact that we find higher project grades also improve average exam scores is important.

The project is a direct application of skills and concepts learned throughout the course via lectures and homework assignments. The exams are built using similar questions to those found on the homework, whereas the project is a comprehensive application of their skills in a different setting, one that is far less controlled and restrictive than their homework and exam setting. We interpret the finding that a higher average project grade improves average exam grades as evidence that the project solidifies learning of course material. Additionally, we determined that the second step of the project that focuses on linear regression analysis is responsible for this enhanced learning of key concepts. We believe this is the case because the first step of the project is repetition of information covered in prerequisite courses.

Overall, our paper contributes to the literature by examining the effects of Project-Based Learning in an applied business statistics course as most studies set in higher education revolve around engineering courses. Our findings provide further support for its effectiveness, extending the literature on Project-Based Learning's impact in a college environment. Ultimately, we find that the project has a positive impact on student learning, as evidenced by higher average exam grades and course grades.

REFERENCES

Al-Balushi, S. M., & Al-Aamri, S. S. (2014). The effect of environmental science projects on students' environmental knowledge and science attitudes. *International Research in Geographical and Environmental Education, 23*(3), 213-227.

Bailey, B., Spence, D. J., & Sinn, R. (2013). Implementation of discovery projects in statistics. *Journal of Statistics Education, 21*(3), 1-24.

Borgaonkar, A. D., Sodhi, J., Vijayabalan, R., & Nair, A. S. (2021). A Comparative Study of the Impact of COVID-19 Pandemic on Student Participation and Performance in First-Year Engineering Courses. *2021 ASEE Virtual Annual Conference Content Access* (pp. 1-9). Virtual Conference: ASEE Conferences. Retrieved from https://peer.asee.org/36566

Chen, C.-H., & Yang, Y.-C. (2019). Revisiting the effects of project-based learning on students' academic achievement: A meta-analysis investigating moderators. *Educational Research Review, 26*, 71-81.

Crossouard, B. (2012). Absent presences: The recognition of social class and gender dimensions within peer assessment interactions. *British Educational Research Journal, 38*(5), 731-748.

Deckard, N. M. (2017). Statistics Education for Undergraduate Sociology Majors: Survey Findings across Institutions. *Numeracy: Advancing Education in Quantitative Literacy, 10*(2), 1-17.

Dierker, L., Evia, J. R., Singer-Freeman, K., Woods, K., Zupkus, J., Arnholt, A., . . . Rose, J. (2018). Project-based learning in introductory statistics: Comparing course experiences and predicting positive outcomes for students from diverse educational settings. *International Journal of Educational Technology and Learning, 3*(2), 52-64.

Efstratia, D. (2014). Experiential education through project based learning. *Procedia-Social and Behavioral Sciences, 152*, 1256-1260.

Fernandes, S. R. (2014). Preparing graduates for professional practice: findings from a case study of Project-based Learning (PBL). *Procedia-Social and Behavioral Sciences, 139*, 219-226.

Formby, S. K., Medlin, D., & Ellington, V. B. (2017). Microsoft Excel: Is it an important job skill for college graduates? *Information Systems Education Journal, 15*(3), 55-63.

Guo, P., Saab, N., Post, L. S., & Admiraal, W. (2020). A review of project-based learning in higher education: Student outcomes and measures. *International Journal of Educational Research, 102*, 101586. doi:https://doi.org/10.1016/j.ijer.2020.101586

Kokotsaki, D., Menzies, V., & Wiggins, A. (2016). Project-based learning: A review of the literature. *Improving Schools, 19*(3), 267-277.

Lee, J. S., Blackwell, S., Drake, J., & Moran, K. A. (2014). Taking a leap of faith: Redefining teaching and learning in higher education through project-based learning. *Interdisciplinary Journal of Problem-Based Learning, 8*(2), 19-34.

Levin, P. (2003). Running group projects: dealing with the free-rider problem. *Planet, 9*(1), 7-8.

Levine, D. M., Stephan, D., & Szabat, K. A. (2021). *Statistics for Managers Using Microsoft Excel.* Pearson.

Oguz-Unver, A., & Arabacioglu, S. (2014). A comparison of inquiry-based learning (IBL), problem-based learning (PBL) and project-based learning (PJBL) in science education. *Academia Journal of Educational Research, 2*(7), 120-128.

Pearson. (2023). *MyLab Statistics*. Retrieved April 21, 2023, from Pearson: https://mlm.pearson.com/northamerica/mystatlab/educators/statistics-features/index.html

Petocz, P., & Reid, A. (2005). Something strange and useless: Service students' conceptions of statistics, learning statistics and using statistics in their future profession. *International Journal of Mathematical Education in Science and Technology, 36*(7), 789-800.

The White House. (2020, March 16). *15 Days to Slow the Spread*. Retrieved April 24, 2023, from Trump White House Archives: https://trumpwhitehouse.archives.gov/articles/15-days-slow-spread/

The White House. (2020, March 24). *Remarks by President Trump, Vice President Pence, and Members of the Coronavirus Task Force in Press Briefing*. Retrieved April 24, 2023, from Trump White House Archives: https://trumpwhitehouse.archives.gov/briefings-statements/remarks-president-trump-vice-president-pence-members-coronavirus-task-force-press-briefing-9/

Wurdinger, S., Haar, J., Hugg, R., & Bezon, J. (2007). A qualitative study using project-based learning in a mainstream middle school. *Improving Schools, 10*(2), 150-161.

APPENDIX

Table A.1
Estimates of Project Impact on Grades – Indicator

	Final Grade Original	Final Grade	Average Exam Grade Original	Average Exam Grade
Project	3.1060** (1.41)	3.6038* (1.96)	-2.5660* (1.43)	1.9747 (2.00)
American Indian or Alaskan Native	9.9305** (5.01)	8.6680* (4.96)	6.6368 (5.08)	5.5657 (5.05)
Asian	9.1351*** (3.23)	9.9313*** (3.21)	8.6596*** (3.28)	9.1433*** (3.26)
Black or African American	-1.5563 (1.21)	-1.5025 (1.20)	-1.9797 (1.22)	-1.9962 (1.22)
Hispanic	1.2598 (0.99)	1.5292 (0.99)	0.6853 (1.01)	0.9256 (1.00)
International	5.0698* (3.00)	5.0795* (2.98)	5.6164* (3.04)	5.8206* (3.03)
Two or More	-0.7047 (1.96)	-0.7715 (1.95)	-1.2686 (1.99)	-1.1864 (1.98)
Unknown or Not Reported	-1.0779 (5.43)	-0.9384 (5.40)	-4.7935 (5.51)	-3.9933 (5.49)
Gender (Female=1)	-0.2182 (0.79)	-0.2793 (0.78)	-1.2977 (0.80)	-1.3576* (0.80)
Instructor (Instructor 1=1)	3.9113*** (0.83)	4.1444*** (0.84)	3.2937*** (0.84)	3.6493*** (0.86)
Time	-0.6101** (0.29)		0.1615 (0.29)	
Spring 2019 (Time=Spring 2019=1)		3.4683* (1.96)		1.0384 (1.99)
Fall 2019 (Time=Fall 2019=1)		5.7878*** (1.98)		4.8225** (2.02)
Spring 2020 (Time=Spring 2020=1)		5.3198*** (1.92)		5.9791*** (1.96)
Fall 2020 (Time=Fall 2020=1)		-0.1610 (1.93)		1.6062 (1.96)
Spring 2021 (Time=Spring 2021=1)		-1.1368 (1.96)		0.9759 (1.99)
Fall 2021 (Time=Fall 2021=1)		-1.8249 (1.52)		-2.8434* (1.55)
Spring 2022 (Time=Spring 2022=1)		0.0000 (0.00)		0.0000 (0.00)
Fall 2022 (Time=Fall 2022=1)		-0.0149 (1.60)		-0.4234 (1.62)
Age	-0.0761 (0.13)	-0.0505 (0.13)	-0.0899 (0.13)	-0.0830 (0.13)
Cumulative GPA	12.4349*** (0.63)	12.5827*** (0.62)	11.9241*** (0.63)	12.0286*** (0.63)
Attempted Hours	0.0038 (0.17)	0.0553 (0.17)	0.0556 (0.18)	0.0846 (0.18)
R^2	0.299	0.317	0.271	0.286
Observations	1202	1202	1202	1202

Note: The dependent variable in the first and second columns is final grade. The dependent variable in the third and fourth columns is the average score on Exams 1-3. The variable of interest is an indicator equal to 1 if the student was subject to the project. Standard errors are reported in parentheses below the coefficients. * p<.10, ** p<.05, *** p<.01.

Table A.2
Estimates of Project Impact on Grades – Average Project Grade

	Final Grade Original	Final Grade	Average Exam Grade Original	Average Exam Grade
Average Project Grade	0.2912*** (0.04)	0.2929*** (0.04)	0.1372** (0.05)	0.1396** (0.05)
Average Peer Evaluation Score	0.2314*** (0.02)	0.2354*** (0.02)	0.2284*** (0.03)	0.2340*** (0.03)
American Indian or Alaskan Native	3.6035 (9.47)	4.5285 (9.44)	5.0783 (12.50)	6.3785 (12.45)
Asian	8.5168** (3.63)	8.7494** (3.62)	10.3425** (4.80)	10.6695** (4.77)
Black or African American	-1.8669 (1.46)	-1.6020 (1.46)	-3.1304 (1.93)	-2.7580 (1.93)
Hispanic	1.7480 (1.16)	1.9067* (1.16)	1.5071 (1.53)	1.7301 (1.53)
International	9.6637** (4.30)	10.1159*** (4.28)	11.9019** (5.67)	12.5376** (5.65)
Two or More	-1.4972 (2.42)	-1.2736 (2.42)	-2.9461 (3.20)	-2.6319 (3.19)
Unknown or Not Reported	6.1652 (6.70)	6.9220 (6.69)	8.9749 (8.85)	10.0387 (8.82)
Gender (Female=1)	0.0192 (0.96)	-0.0687 (0.96)	-0.3436 (1.27)	-0.4671 (1.26)
Instructor (Instructor 1=1)	3.6446*** (0.92)	3.6226*** (0.92)	3.8343*** (1.22)	3.8033*** (1.21)
Time	0.1511 (0.57)		0.6018 (0.75)	
Spring 2022 (Time=Spring 2022=1)		2.1708** (1.10)		3.4408** (1.45)
Fall 2022 (Time=Fall 2022=1)		0.1445 (1.13)		0.9819 (1.49)
Age	-0.1974 (0.14)	-0.2104 (0.13)	-0.2571 (0.18)	-0.2755 (0.18)
Cumulative GPA	7.3503*** (0.72)	7.4465*** (0.72)	8.9932*** (0.95)	9.1285*** (0.95)
Attempted Hours	0.0487 (0.21)	0.0267 (0.21)	0.1707 (0.27)	0.1398 (0.27)
R^2	0.562	0.567	0.428	0.435
Observations	428	428	428	428

Note: The dependent variable in the first and second columns is final grade. The dependent variable in the third and fourth columns is the average score on Exams 1-3. The variable of interest is the average project grade across two steps. Estimates are conditional on the student being subject to the project. Standard errors are reported in parentheses below the coefficients. * p<.10, ** p<.05, *** p<.01.

Table A.3
Estimates of Project Impact on Grades – Step 1 Project Grade

	Exam 2 Original	Exam 2	Exam 3 Original	Exam 3
Step 1 Project Grade	0.1238 (0.08)	0.1236 (0.08)	0.0421 (0.08)	0.0411 (0.08)
Average Peer Evaluation Score	0.2199*** (0.04)	0.2218*** (0.04)	0.3258*** (0.04)	0.3352*** (0.04)
American Indian or Alaskan Native	-5.8794 (17.46)	-5.4410 (17.49)	12.6036 (17.96)	14.7279 (17.86)
Asian	12.2485* (6.70)	12.3612* (6.70)	9.5984 (6.89)	10.1446 (6.85)
Black or African American	-4.2053 (2.70)	-4.0822 (2.71)	-1.9804 (2.77)	-1.3839 (2.76)
Hispanic	1.7184 (2.14)	1.7945 (2.15)	1.7625 (2.20)	2.1315 (2.19)
International	13.4508* (7.92)	13.6723* (7.93)	13.2528 (8.14)	14.3260* (8.10)
Two or More	-0.9526 (4.47)	-0.8504 (4.47)	-7.8620* (4.59)	-7.3668 (4.57)
Unknown or Not Reported	13.2556 (12.35)	13.6049 (12.37)	1.2321 (12.70)	2.9247 (12.63)
Gender (Female=1)	-2.9606* (1.77)	-2.9987* (1.77)	1.6975 (1.82)	1.5128 (1.81)
Instructor (Instructor 1=1)	8.3218*** (1.70)	8.3092*** (1.70)	3.9130** (1.75)	3.8519** (1.73)
Time	1.1530 (1.04)		0.8430 (1.07)	
Spring 2022 (Time=Spring 2022=1)		2.1030 (2.04)		5.4470*** (2.09)
Fall 2022 (Time=Fall 2022=1)		2.2350 (2.09)		1.3424 (2.14)
Age	-0.0670 (0.25)	-0.0732 (0.25)	-0.5501** (0.26)	-0.5801** (0.26)
Cumulative GPA	9.2486*** (1.32)	9.2960*** (1.33)	10.5770*** (1.36)	10.8069*** (1.35)
Attempted Hours	0.3723 (0.38)	0.3622 (0.38)	0.3774 (0.39)	0.3285 (0.39)
R^2	0.299	0.300	0.363	0.373
Observations	428	428	428	428

Note: The dependent variables are Exam 2 in the first and second columns and Exam 3 in the third and fourth columns, which occurred after Step 1 of the project was due. The variable of interest is the Step 1 project grade. Estimates are conditional on the student being subject to the project. Standard errors are reported in parentheses below the coefficients. * p<.10, ** p<.05, *** p<.01.