CASE STUDY

An analysis of student reflections of semester projects in introductory statistics courses

Richard Spindler, Department of Mathematics, SUNY Plattsburgh, Plattsburgh NY, USA. Email: rspin004@plattsburgh.edu

Abstract

Projects in an introductory statistics course have been implemented for many years as an authentic assessment and learning activity. This article examines student perceptions toward semester projects and the relationship of those perceptions to outcomes. Student semester project data including personal reflections and assessment scores from 254 undergraduate students in six semesters of introductory statistics were analysed using qualitative and quantitative methods, relying heavily on content analysis. The projects comprehensively assessed introductory statistics content, and student reflections were based on three prompts. High levels of learning and interest were reported by the students. Workload issues were not expressed by students to an appreciable degree. Finally, this study illustrates the value of using qualitative methods with case studies to inform teaching and learning.

Keywords: Statistics, Teaching, Project, Qualitative Method

1. Introduction and Literature Review

Projects and other authentic assessments, defined by directly examining student performance on worthy intellectual tasks (Wiggins, 1990), have been proposed as important tools to not only monitor but improve performance, paraphrased from Wiggins (1990). There are many examples and case studies of authentic tasks available. For a few recent examples, Shekar (2014) describes a re-design of an engineering curriculum around project-based learning, Marshall (2019) discusses how statistical education guidelines were used to change a first year probability and statistics module into a project based learning approach, and Spindler (2019) describes modelling projects in differential equations and a rubric for assessing them. For many years the statistical education community has expressed concern about the focus on learning formulas and techniques in introductory statistics courses, instead of on statistical thinking and the statistical process of collecting data to answer questions about the 'real' world (e.g., Garfield, 1994; Cobb and Moore, 1997; Cobb 2007; American Statistical Association, 2016). In addition, in any field student assessment test anxiety can be a concern and performance assessments may lower that anxiety (Onwuegbuzie, 2000).

Students' perceptions about assessment are correlated with what and how they learn. In a comprehensive review of the literature of students' perceptions about assessment, Struyven, Dochy, and Janssens (2005) reveal that the type of assessment influenced the type of learning that students engaged in (i.e. shallow versus surface learning). Thus, it is important to understand students' perceptions of their work. Using case studies with interviews, Sambell, McDowell, and Brown (1997) studied students' perceptions of alternative assessments in a wide variety of subject areas. They reported that students viewed alternative assessments to tests as more authentic compared to the artificial nature of traditional assessment, as a more accurate measure of learning compared to traditional assessment, as measuring complex skills (compared to memorization), and to be fairer (closely related to impressions of validity) than traditional tests. Similarly, Onwuegbuzie (2000) also reported that students rated authentic statistics assessments as promoting higher-level thinking compared to tests. Early (2007) used a phenomenological approach with student interviews to better understand student experiences in introductory statistics courses. Likewise, using qualitative methods, Neumann, Hood, and Neumann (2013) studied students' perceptions of using real-life data

in introductory statistics course instruction and identified authenticity as being positively associated with student learning experiences. Finally, Moreira and Pinto (2014) created an end-of-course evaluation that students completed at the end of a 'learning project', which they describe as *"a dialogical teaching approach which can remove emotional inhibitions of students who are fearful of calculations involved in statistics and encourage them to interact among themselves and learn via a team approach"* (Moreira and Pinto, 2014, p. 178). The data from the end-of-course evaluation included both quantitative and qualitative information, and they concluded that the learning projects help students connect statistical concepts and practical applications.

The present study builds on the work of these findings on students' perceptions of alternative assessments by studying students' reflections on a semester project for 254 students in an introductory statistics course and examining different aspects of their reflections. In addition, this study provides a case study of how qualitative methods may inform instructors on their teaching and on student thinking. The specific intent of this retrospective study is to explore students' perspectives of an introductory statistics semester project and the impact of the semester projects on student learning for students at primarily 4 year public universities. These projects were authentic assessments since students were analysing real data and whose goal was to reach conclusions based on students applying appropriate statistical analyses. Student project reflections were studied both through content analysis and quantitative methods on prior known data. Prior known data included final exam and project scores, project types, and the semester when the projects were completed.

The primary goal was to explore and identify undergraduate student impressions of the project and to explore relationships between these impressions and outcomes. The central questions explored are as follows. What did students feel or express about the projects and the value (or not) of the projects? What themes or categories of student sentiment emerge? How are those categories of student expressions related to each other and the known prior outcomes of the course, especially project scores? Philosophically, the approach here is a practical one and could be considered *interpretive description* which is *"to discover association, relationships, and patterns within the phenomena that has been described"* (Thorne, 2016, p. 56).

2. Method

2.1 Participants

The participants of this retrospective study were undergraduate students in an introductory statistics course where a semester project assessment was used. The students for the first five semesters of this study were at a primarily undergraduate traditional mid-western regional public university consisting of approximately 10,000 students with a strong liberal arts focus but also with substantial professional programs such as nursing and business. The sixth and last semester students were similar, attending a primarily undergraduate eastern regional public university consisting of approximately 6,000 students again with liberal arts as an important focus along with some professional programs. Each section typically consisted of 30-35 students. The students were a diverse set of majors, some requiring statistics and some meeting just general education requirements. Since this is a retrospective study, students were not randomly assigned to the introductory statistics sections. However, all introductory statistics sections were listed in the course catalogue in the same way and so the section using projects were not identified as such.

2.2 Course Background

All of the students in this study were taught by one instructor (the author), and I had taught introductory statistics for many semesters. Prior to the Spring of 2014, over a few semesters I had shifted the course from a standard lecture-based course to a flipped classroom model. Students

would read or watch videos before class. Then the students worked in groups on learning activities and problems of varying types in class, coupled with class discussions and mini-lectures.

For assessment, during that transition time period a major semester project was instituted to be due at the end of the semester. To ensure that students received feedback on their understanding of the content and to ensure accountability, midterm exams were still part of the course, but their weight on the final grade was reduced. The students still took a mandatory, department-wide common final exam, but its weight on the final grade was small compared to the semester project. The semester projects were instituted prior to Spring of 2014, but those data were not included in this study since important changes during those semesters were made to the instructional approach and the assessment of the projects as I learned and adapted. By the Spring of 2014, the course instruction and the project were stable.

In the semester project, I encouraged students to work in pairs but they could choose to work alone. They could choose between utilizing authentic data provided to them, or creating a survey of their own and gathering the data, which were called personal projects. In the former, I provided National Survey of Student Engagement (NSSE) data for the institution. In addition, I obtained real data about food insecurity in the community gathered in prior years by a community research team independent of the course. Thus, students had a choice of three project types: personal, NSSE, or food insecurity. For the initial segment of the project, students chose survey questions from the surveys used to gather the data or created their own if doing a personal project. In addition, they created research questions based on those survey questions. Note the process was not necessarily sequential this way, as creating the research questions and choosing/creating survey questions were dependent on each other. This was typically a productive learning process for the students to understand the difference between research questions and survey questions. I emphasized the importance of this part of the project by receiving two submittals to review to ensure the students would have solid research and survey questions to work with. In addition, students were required to submit a basic timeline with dates for finishing project milestones, which included a written introduction, a written methods section, data acquisition (if a personal project), descriptive statistics analysis, inferential statistics analysis, and a written conclusion with reflection.

The heart of the project was the analysis. In the final project paper submitted at the end of the semester, students were required to answer their research questions using both exploratory (descriptive) and inferential statistics from the data. Once students learned enough procedures, either descriptive or inferential, in class they were guided by activities to help them decide what procedure made sense with what type of research question. Although working through the procedures was important, a more critical part of their work was interpreting and evaluating their results, justifying if a procedure used made sense and was consistent with other results. Throughout the semester, in class students would learn, practice, and elaborate on those elements. When students had learned enough descriptive statistics content in the course, I required students to submit draft work on the descriptive statistics portion of their analysis to help with time management and to benefit from some feedback. This helped students think about time management and planning during the semester. In addition, as the deadline for the project approached, I provided time in the classroom for students to work together and to gain feedback on their progress.

A detailed rubric with specific guidelines was used to guide the students and for assessment. Suskie (2004) gives an overview of assessment tools and in particular rubrics. Menendez-Valera and Gregori-Giralt (2015) cites the literature on the many potential benefits of using rubrics, which include setting learning standards, allowing students to make academic decisions, giving students tools to assess their own progress, making grading more transparent, providing students feedback on their work, and enhancing communication between faculty and students focused on learning. Similarly, Chance (1997) stresses the importance of providing students ‘guidelines of what is expected’.

Transparency is an essential part of using rubrics (Jonsson, 2014), and so activities on interpretation and evaluation in class helped students understand the semester project rubric. Thus, a rubric provides important support for learning beyond grading.

2.3 Data Collection

The project rubric included a short component for the students to reflect on the project experience. The reflection prompts given to the students were:

- Did you encounter any problems?
- Did other questions arise? Identify interesting follow-up investigations.
- Reflect on the project experience as a whole from a personal point of view. For example, did you learn from it? Did you enjoy working with real data? Was it meaningful?

The purpose of these reflections was for students to think about what they gained from the project experience (or not). The data analysed in this study are the reflections based on those prompts, semester project scores, final exam scores, and type of project over six semesters from Spring, 2014 to Fall 2016, for each student. At the time, I had no thought about using their reflections for analysis, so the study is retrospective. (Since then, I did receive university IRB approval to analyse the data.) Data from a few students that did not write a reflection and students that did not take the final exam were eliminated before analysis. This resulted in reflection data for 254 students. Students that worked as partners could write separate reflections or write one reflection together if they both agreed to the content of the reflection.

2.4 Data Analysis

I used both qualitative and quantitative analysis techniques to analyse the data from these 254 students. Partners that wrote the same reflection were counted twice since the implication is that both felt the same way as reported in that reflection. The text of the reflections were analysed qualitatively using an inductive coding approach as described in the grounded theory method (Babbie, 2014, p. 405). However, it should be noted that theoretical saturation was not used because this study analysed fixed, historical data. Thus, the qualitative aspect of this study was not a true grounded theory approach but primarily used coding ideas from grounded theory. Of course, because most of the information comes from a content analysis of the reflection data, categories or themes were not known ahead of time. However, data that were available prior to the analysis were project scores, final exam scores, project type, and semester. Although the original student words and sentiments summarized in the quotes and content analysis are the most meaningful, I also used quantitative methods to analyse both emergent categories and prior data, and also to analyse relationships between these.

The reflections were coded in an iterative process by beginning with open coding and then continuing with focused (or selective) coding (Glaser, 1978). Thus, first an initial reading and review of the reflections indicated some possible categories or themes that appeared regularly in the reflections. Tentative categories were created based on that initial review. Then, a more thorough reading of the reflections was done with coding and categorizing. This round of focused coding caused a refinement of categories, including re-naming and combining categories, as well as creating sub-categories to parse out finer variations of the categories. This process was simplified by using an electronic spreadsheet, placing all reflections in one column and categories (or sub-categories) in new columns as they emerged from the data. Because the coding was done over multiple sessions, it was important to return to previously coded reflections to understand the meaning of those categories before going forward again.

3. Results

3.1 Categories from the content analysis

Several major categories emerged from the content analysis: Learning, Interest, Time Management/Organization, and Challenge. In addition, a few other less common categories included Partners and Confidence. Table 1 summarizes sub-categories that describe each of these categories.

Table 1. Categories resulting from coding student reflections

Category	**Description through sub-categories**
Learning (L)	Overall learning, Application, Comprehensive Coverage, Concepts, Particular topics (in the student project).
Interest (I)	General Interest, Enjoyed applying, Future interests, Fulfilling, Liked Options, Satisfied/Proud.
Time Management /Organization (T)	Good Time Management, Learned about Time Management, Poor Time Management, Put More Time into.
Challenge (C)	Workload (Much work/Time-consuming), Other challenges.

The Learning category indicated whether a student expressed learning gains and, if possible to identify, in what way. The sub-categories that emerged were general overall learning, applying statistics to real life, comprehensively covering the course, reviewing concepts, and learning about the topic of the student's particular project. Table 2 in the Appendix contains multiple examples of such quotes. Here is one:

"I think it is an excellent educational tool that allows students to not only understand elementary statistics, but also to practically utilize those skills."

In the Interest category, students expressed their interest in the project or other similar positive emotions, such as fulfilment or satisfaction. The sub-categories here included general statements of interest, enjoying applying statistics to real problems, creating future interest in statistical problems, fulfilling work, liking the ability to choose between options for the project, and being proud of the work. Here is one example from Table 3 in the Appendix:

"This project was very interesting overall!"

The Time Management category indicated a comment about time management and organization. Some students explicitly reported having good time management (no problems with time management or managing their time well). Some students reported having some issues with time management. A subset of these latter students reported learning from the experience (expressing positive reactions to it), while some were negative about it, for example blaming the instructor. Many students were silent about time management. Thus, the subcategories include statements indicating time was managed well, learning about managing time, poorly managing time, and wishing more time as put into it (either out of class or in class). Table 4 in the Appendix contains examples, such as

"I did not realize how big this project actually was until I started it and realized I was not going to have enough time to finish it by the deadline. I definitely should have followed the schedule provided for me, so I would have finished the project on time."

Finally, in the Challenge category students expressed the project as challenging in some way with a general sub-category and a workload (hard work or time consuming) sub-category. Table 5 in the Appendix provides examples, including:

"It was a difficult and time consuming project."

3.2 Quantitative analysis of categories

Over three-quarters of the students, 78.3% (199 out of 254), stated some kind of learning benefit, which makes sense since one of the reflection prompts suggested commenting about learning. The largest number of comments were on the great overall learning experience or learning how to apply statistics, with the fewest commenting about learning or reviewing concepts. This also makes sense since the project was focused on analysis, applications, interpretation, and evaluation of results while exams were used to test statistical procedural knowledge and some understanding of concepts. Approximately a third of the students reported learning in more than one way (two or more learning subcategories). Also, over two-thirds of the students, 67.7% (172 out of 254), stated some kind of interest characteristic, with most students stating they enjoyed or were interested in the project and many stating it was interesting to apply the statistics. Almost a quarter stated two or more interest subcategories.

Most of the students did not make a comment about time management. Of the 254 students, only 47 (18.5%) mentioned time management in any form. To break it down further, 15 (5.9%) of all students expressed they had good time management (managed their time well or had no problems with time management), while 32 (12.6%) of all students stated they had issues with time management. Of those 32 students, 16 expressed a negative statement about time management. The remaining of the 32 students expressed learning about time management from the project experience. Finally, 76 (30%) of all students stated it was challenging in some way (though not necessarily in a bad way), while 31 (12.2%) of all students specifically stated it was hard work or time consuming.

3.3 Assessment Score Analysis

Students did reasonably well on projects, especially compared to the common final exam. The median project score was 84% (mean 78%) yet the median final exam score was 71% (mean 67%). In addition, the variability of the project scores (standard deviation 18 percentage points) was lower than for final exams (standard deviation 21 percentage points). The relationship between project scores and final exam scores was also examined, illustrated in the scatterplot in Figure 1. Pearson's linear correlation coefficient for this relationship is 0.39. There is not a clear linear relationship, but it appears reasonable to conclude that the relationship is positive. When the scatterplot is broken down by project type, the resulting three scatterplots are not appreciably different from Figure 1.

Figure 1. Scatterplot of final exam scores versus project scores for all 254 students.

3.4 Relationships with project scores

Project scores did appear to be related to students' expression of learning, since the median score when no learning was reported was 75 (mean 74.1), and the median score when learning was expressed was 86 (mean 83.8), also illustrated in Figure 2. A Mann-Whitney test yielded a p-value of less than 0.001.

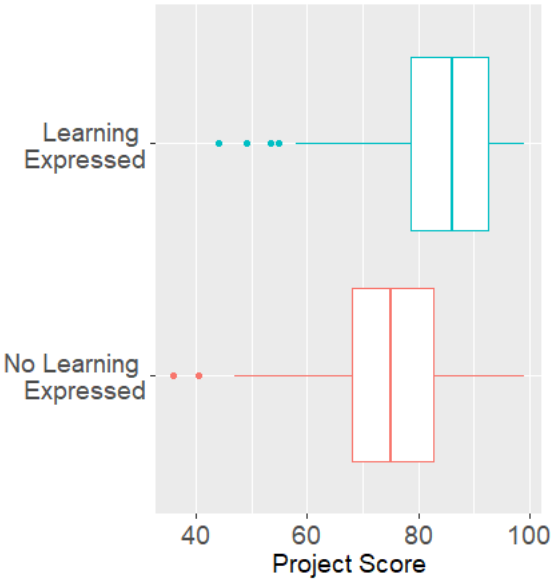

Figure 2. Project Scores by whether learning was reported. Out of the 254 students, 199 expressed learning from the project in at least one way.

The 15 students in the sub-category who explicitly said they had good time management appeared to do appreciably better on the project. The median score when good time management was reported was 90.5 (mean 86.0) and for all other students the median score was 84.0 (mean 81.5). However, the small sample and a borderline p-value of 0.05 with the Wilcox-Mann test implies a questionable significance. The relationship between time management and project score is

supported somewhat by combining students that expressed they had good time management with students that had problems with time management but were positive about learning time management. There were 31 students that expressed either trait. The median score for students that expressed good time management or were positive about learning time management was 90.0 (mean 85.2) and for all other students the median score was 83.5 (mean 81.2), with a Wilcox-Mann test p-value of 0.02.

Project score did not appear to be affected by students' reporting the work being challenging in the Challenge category. However, within that category, some students (31) specifically said the project was hard work or time consuming. It was found that there may be some relationship between the project score and if the student specifically expressed the project being hard work or time consuming (Wilcox-Mann p-value of 0.05). The median score when hard work was reported was 88.0 (mean 85.0) and for all other students the median score was 83.3 (mean 81.3).

Project scores did appear to be affected by whether students had partners (Wilcox-Mann p-value less than 0.001). The median score of students with a partner was 86.0 (mean 84.7) and for students without a partner, the median score was 81.3 (mean 77.7). In addition, the variability of scores was appreciably lower for students with partners (standard deviation 9) versus for students without partners (standard deviation 15). Figure 3 illuminates the results about partners further, also showing the left skew of project scores for students without partners.

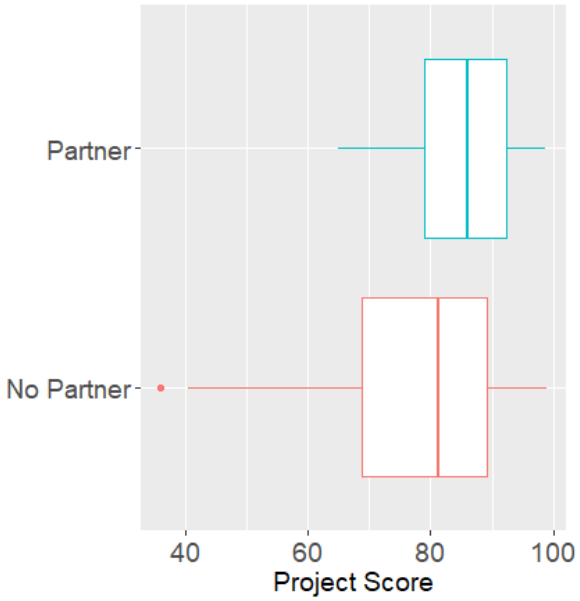

Figure 3: Project Scores by whether a student had a partner. Out of the 254 students, 107 did not have a partner. (Note that an odd number of students with a partner in the data occurred because one student with a partner did not write a reflection and so his/her data was dropped.)

3.5 Relationships with expressions of learning

The possible dependence of students' expression of learning on other factors is examined next, starting with time management. Recall some students reported good time management (having no problems with time management or managing their time well), some students reported having some issues with time management (a subset of these latter students reported learning from the experience, while some were negative about it), and some students were silent about time management. Again, note the low percentage, 18.5%, that expressed any sentiment about time management. A comparison of the learning expressed between students that reported a problem with time management and were negative about it (n=16) compared to all other students indicated

that a negative attitude about time management may be associated with a lower expression of learning, but the numbers are too small to make any conclusions, confirmed using a chi-square test. In addition, unlike project scores, there was little difference in expressed learning between students that were positive about time management (good time management or learning from it) and the rest of the students.

Similarly to project scores, a student expression of learning does appear to be related to having a partner (Chi-squared test p-value of 0.003). Of students with a partner, 85% expressed learning, while of students without a partner, 69.2% expressed learning. Finally, there did not seem to be an appreciable difference in scores, learning, interest, or challenges across semesters or across the three types of projects.

4. Discussion

The analysis above documents and quantifies student expression and student outcomes on an authentic assessment. Students overall viewed the semester statistics project as a comprehensive, useful, and interesting learning experience, as illustrated by the quotes in Table 2 in the Appendix on comprehensive coverage and learning applications. In addition, a student's expression of learning was highly related to the project score. There could be a number of reasons for this relationship, but it seems reasonable to conjecture that students who think they are learning much from the project, aware of its importance in their learning, will try to do good work. This observation is related to the concept of self-regulated learning (SRL). Self-regulated is 'the degree that they [students] are meta-cognitively, motivationally, and behaviourally active participants in their own learning process' (Zimmerman, 1989). In this case, a reasonable conjecture is that if motivation related to learning is high, then outcomes will improve.

It is possible that the third reflection prompt may have implicitly biased students to state that they learned from the experience. In part of that prompt, students were asked: 'Reflect on the project experience as a whole from a personal point of view. For example, did you learn from it? ….' Thus, perhaps more diligent or ambitious students would make a point of mentioning learning and also those same students would do well on the project. However, the extent of some of the comments, stating not only whether they learned but what they learned (with a third of the students stating more than one subcategory), and the intensity of the student comments on learning, appear to indicate that there is more to those reports of learning than to meet a reflection component of the rubric or to please the instructor. For example, students genuinely expressed a range of areas of what they learned, including how to apply the content, comprehensive coverage of the material, deeper understanding of statistical concepts, and learning particular topics (such as about food insecurity), as illustrated in Table 2 in the Appendix. The reflection also had a low impact on their project score and saying something specifically about learning was not required.

Students also reported a high degree of interest in the project, as illustrated in Table 3 in the Appendix and in the content analysis. One of the prompts was *"Did you enjoy working with real data?"* which again may have biased student reflections. But, again, the depth and variety of the comments (with a quarter of the students expressing more than one category) seem to reveal genuine interest rather than to gain points on a low impact part of the project. It would be interesting to apply more formal attitude tools, such as described by Gal and Ginsburg (1994), to study the role of interest and beliefs about learning in determining outcomes in the context of semester statistics projects.

A positive result was how well students overall performed on projects compared to final exams, and more consistently as shown by the lower project score variability. There could be many factors related to these results, including that projects don't measure the same outcomes and learning characteristics as final exams. This is partly true in that the projects assessment is more inclusive.

The final exams focused mostly on procedures and some understandings of statistical and probability concepts, while the projects included application, interpretation, and evaluation, along with the procedures and some analysis. It is also possible that students approached a project more seriously than a final exam since the project was worth appreciably more than the final exam as a percentage of the final grade. However, the median score of the final exams for five sections over the several years prior to the use of the projects (with the same instructor) was a 74 (mean 73). This is still 10 points lower than the median project score. In other words, students did better on projects even when compared to students that had only a final exam as the last major course assessment (worth about 25% of the grade).

An expectation I had entering into this study was that a high proportion of students would say the semester project was a lot of work and time consuming. But the data did not show that. This conclusion is reinforced in that a relatively small percentage mentioned time management, good or bad. Although this seems counterintuitive, Kember (2014) examined case studies revealing that students' impression of workload is not equivalent to the time actually spent on tasks. He found that some factors that influence and correlate with a lower impression of workload include concentrating on concepts and understanding, active engagement of students ['projects seem to be particularly effective', (Kember, 2014, p. 182)], positive student-student-teacher relationships, and class coherence. All of these were generally present in this course.

On the other hand, students that reported the project to be hard work or time consuming tended to perform somewhat better on the projects than students that did not, although the difference was not large and the sample size of 31 is relatively small. One conjecture is that the expression of hard work may be an indication of persistence and how seriously the student took the project. Also, time management is important for a potentially time-consuming project, and the results show that. The data on time management, though not definitive and from a small sample, indicate that a student's approach to time management plays a role in success as measured by project scores. This is also supported by the results indicating that poor time management may be related to lower expressions of learning.

Having a partner did appear to improve student project scores, where students with partners scored better on projects and tended to express learning. These results seem to reflect the idea that teamwork and cooperative learning supports student learning and outcomes (Springer, Stanne, and Donovan, 1999; Gaudet, et al, 2010; Slavin, 2014). In addition, not having a partner appeared to produce higher variability in student scores. The primary source of this higher variability was a left skew, where 25% of non-partnered students had a project score below 69%. In their meta-analysis study, Springer, Stanne, and Donovan (1999) describe the positive motivational effects of group learning in science, mathematics, engineering, and technology. A semester project is a complex and intensive undertaking, and so it may be that some of the non-partnered students did not have the persistence and mutual support that is available to partnered students to finish the project completely and strongly. It would be interesting to explore this further to determine if this subset of students has particular characteristics, for example, under-represented groups, traditionally less college-prepared students (such as first generation students), or other characteristics impeding their reception to group work.

Finally, the type of semester project did not appear to be associated with outcomes and other factors. Thus, students creating their own survey and gathering their own data did not appear to influence their outcomes differently when compared to being provided authentic data that they could analyse instead. Thus, it may be that simply having interesting and authentic data regardless of the source is the important factor.

Because this study was done retrospectively and the data is mostly qualitative in nature, the conclusions made here are necessarily exploratory and tentative. In addition, the participants were

not chosen randomly. However, one strength of this study is that several important factors were kept constant: one experienced instructor throughout, a consistent course curriculum and pedagogy in the classroom, the same assessment tool of the projects through a rubric, and a mostly consistent department-wide final exam. Except one semester, all students were attending the same university, and both universities were similar in purpose and types of students they served. The courses were always taught Tuesday and Thursdays during the day. The lack of differences in project and final exam scores across semesters is likely a consequence of the consistency in those factors. Another strength is that there is a significant number of participants. The student reflection questions were open-ended which can be both a strength and a weakness. As discussed above, there may have been implicit bias in these questions. Also, students expressed what is most important to them, yet every student did not express thoughts on every category. So, students could have felt a certain way on a characteristic but just didn't write it. The assumption was made that a student didn't feel strongly about that particular characteristic if the student did not report on it.

5. Conclusion

To conclude, this retrospective exploratory study utilized qualitative and quantitative methods to analyse student reflections by 254 students on semester statistics projects. It appears that many students working on projects believe they make important gains in learning on how to use real data within the big picture of statistical analysis and also many appear interested in such projects. Students generally performed better on a project than on a final exam. In addition, student expression of learning gains appeared to be strongly associated with improved project scores. Perhaps surprisingly, a low percentage of students reported problems with having enough time to complete the semester project. This study also supports the notion that cooperative learning, at least through partners, can be beneficial to student learning. Finally, these results suggest the value of future qualitative studies and case studies examining ways for students to be given assessments and activities where students feel they are having significant learning experiences and are actively engaged in authentic activities.

6. Acknowledgements

I am grateful for the advice and manuscript review generously given by Dr. Elizabeth Onasch at SUNY-Plattsburgh. In addition, I want to thank Rosa Canales for assistance in reviewing and analysing the student reflections.

7. Appendix

Table 2. Student examples of the Learning Category

Description through sub-categories	Quotes
L: Overall learning	*"Doing this also taught me to figure out when and how to use each test in different scenarios without a book telling me to use a specific test or graph to test and show my data."* *"I think it is an excellent educational tool that allows students to not only understand elementary statistics, but also to practically utilize those skills."*
L: Application	*"It was a good experience working with a project that had real data and was local."*

	"This was the most hands-on learning I can recall participating on and I feel that is how I am better suited to learn."
L: Comprehensive Coverage	*"This project served as a good review for inferential stats because we had to use many different types of tests. We had to do quite a bit of reviewing."* *"As a whole, I think this project is a great way to end the semester because it tied in everything we learned throughout the entirety of this course. I remember learning about qualitative and quantitative variables the first day of class."*
L: Concepts	*"It helped us understand the concepts within the class much more thoroughly."* *"Not only do I recognize the concepts I learned in class, but I also understand what those concepts measure and how they apply to answering a problem and forming a conclusion."*
L: Particular topics (in the student project)	*"This project taught both of us a lot about food insecurity. We now we realize that this is a growing problem and is more wide-spread than just large urban populations."* *"We thought it was super interesting finding out the different answers that we got and getting an idea of what the population of active individuals is doing with their workouts."*

Table 3. Student examples of the Interest Category

Description through sub-categories	Quotes
I: General Interest	*"This project was very interesting overall!"* *"The project as a whole was very engaging."* *"Overall, the project was a very interesting topic, especially when applied to our area right here."* *"Overall we enjoyed the experience since we were able to work at a pace that was more flexible toward our different schedules."* *"It was cool to put all of the information we had been learning over the semester into application and see what could happen."*
I: Enjoyed applying	*"So this has been my favorite final project in a college class since it has real world application and really is a*

	good representation of all that we have learned throughout the semester." *"It was enjoyable to research something that students deal with on a day to day basis and to use what we have learned to conclude the things that we did."*
I: Future interests	*"This project hit right on the spot for what I want to do for a career making it very interesting to me and made me want to learn more and do more research on the topic as to how I can help now and later on."*
I: Fulfilling	*"I enjoyed this challenging but fulfilling project!"* *"I think that I definitely grew from this project because it pushed me to really work hard and to look at connections from data, which I normally would not do if I hadn't been assigned to do so in class. As a nursing student, I don't really get to spend too much time making calculations and analyzing data too often, so I enjoyed this challenging but fulfilling project!"* *"Thank you for making me do this project and forcing me to realize some pretty awesome things."*
I: Like Options	*"We are glad that we had the option of doing the food insecurities project because it was really eye opening for us to see that the city of Eau Claire isn't as perfect as we thought."* *"We also appreciated the fact that we got to choose our own subject of research and also chose our options of how we would like to analyze our data."*
I: Satisfied/Proud	*"We are extremely happy with this project and our results."* *"I was proud of myself"* *"We are proud of our work."*

Table 4. Student examples of the Time Management Category

Description through sub-categories	Quotes
T: Good Time Management	*"We knew it was going to be a process and take quite a bit of time, so we met at least once most weeks throughout the semester. Because we managed our time well, we were able to finish on time with no struggles or rushing."*

	"This project wasn't too much of a workload as long as we stayed on top of it, which we did a pretty good job on."
T: Learned about Time Management or admitted to poor Time Management	*"I should have made sure that I managed my time better when trying to complete these end of the semester projects and papers. I should have made myself a schedule that I would work on a couple pages a day to complete it. I was able to complete this project, but I could of spread out the work more equally throughout many days."* *"I did not realize how big this project actually was until I started it and realized I was not going to have enough time to finish it by the deadline. I definitely should have followed the schedule provided for me, so I would have finished the project on time."* *"If I could give one piece of advice someone who is just starting this assignment, it would be create a schedule and stick to it."* *"I truly wish I had stuck to my timeline more closely."*

Table 5. Student examples of the Challenge Category

Description through sub-categories	Quotes
C: Hard work/Time-consuming	*"I really enjoyed this project, it was just a lot of work. Which is not necessarily bad if students keep on the pace you set for them."* *"It was a difficult and time consuming project."* *"This project has truly pushed both of us to new limits. This was a very challenging project."*
C: Other challenges	*"This project was not easy by any means, and took a lot of time and dedication and critical thinking."* *"Minitab was very difficult to use at times."* *"Personal things in my life also made problems arise. It was very hard to focus on school in general for a while and I put things on the back burner so that I could overcome the obstacles, which I am still working on."*

Quotes about confidence:

"We feel confident in our research and conclusions."

"I will definitely take away a lot of confidence and understanding of what we have covered all semester."

Quotes about partners:

"Doing it on my own also made it more difficult because I didn't discuss questions and conclusions I made with other people, preventing me from organizing my project and thoughts as well as I could of."

"I enjoyed working with [name removed], and am very grateful that we each pulled our own weight. This project would not have been completed as thoroughly and on time, if we didn't work together as well as we did."

8. References

American Statistical Association, 2016. *Guidelines for Assessment and Instruction in Statistics Education, (GAISE) College Report*. Available at http://www.amstat.org/asa/files/pdfs/GAISE/GaiseCollege_Full.pdf [Accessed 10 January 2020].

Babbie, E., 2011. *The Basics of Social Research (6th Ed.)*. Belmont CA: Wadsworth.

Chance, B., 1997. Experiences with Authentic Assessment Techniques in an Introductory Statistics Course. *Journal of Statistics Education*, 5(3). Available at http://www.amstat.org/publications/jse/v5n3/chance.html [Accessed 10 January 2020].

Cobb, G. and Moore, D. 1997. Mathematics, Statistics, and Teaching. *The American Mathematical Monthly*, 104(9), pp.801-823.

Cobb, G., 2007. The Introductory Statistics Course: A Ptolemaic Curriculum. *Technology Innovations in Statistics Education*, 1(1). Available at https://escholarship.org/uc/item/6hb3k0nz [Accessed 10 January 2020].

Early, M., 2007. Students' Expectations of Introductory Statistics Instructors. *Statistics Education Research Journal*, 6(1), pp.51-66.

Gal, I. and Ginsburg, L., 1994. The Role of Beliefs and Attitudes in Learning Statistics: Towards an Assessment Framework. *Journal of Statistics Education*, 2(2). Available at http://ww2.amstat.org/publications/jse/v2n2/gal.html [Accessed 10 January 2020].

Garfield, J., 1994. Beyond Testing and Grading: Using Assessment to Improve Student Learning. *Journal of Statistics Education*, 2(1). Available at http://www.amstat.org/publications/jse/v2n1/garfield.html [Accessed 10 January 2020].

Glaser, B., 1978. *Theoretical Sensitivity*. Mill Valley. CA: Sociology Press.

Jonsson, A., 2014. Rubrics as a way of providing transparency in assessment. *Assessment & Evaluation in Higher Education*, 39(7), pp.840-852.

Marshall, E., 2019. Embedding and assessing projected based statistics. *MSOR Connections*, 17(2), pp.75-82.

Menendez-Varela, J. and Gregori-Giralt, E., 2015. The contribution of rubrics to the validity of performance assessment: a study of the conservation-restoration and design undergraduate degrees. *Assessment & Evaluation in Higher Education*, 41(2), pp.228-244.

Moreira da Silva, M. and Pinto, S., 2014. Teaching Statistics Through Learning Projects. *Statistics Education Research Journal*, 13(2), pp.177-186. Available at https://iase-web.org/documents/SERJ/SERJ13(2)_daSilva.pdf [Accessed 10 January 2020].

Neumann, D., Hood, M. and Heumann, M., 2013. Using Real-Life Data When Teaching Statistics: Student Perceptions of this Strategy in an Introductory Statistics Course. *Statistics Education Research Journal*, 12(2), pp.59-70.

Onwueguzie, A., 2000. Attitudes toward Statistics Assessments. *Assessment & Evaluation in Higher Education*, 24(4), pp.321-339.

Sambell, K., McDowell, L. and Brown. S., 1997. 'But is it fair?' An Exploratory Study of Student Perceptions of the Consequential Validity of Assessment. *Studies in Educational Evaluation*, 23(4), pp.349-371.

Shekar, A., 2014. Project based Learning in Engineering Design Education: Sharing Best Practices. *ASEE Annual Conference & Exposition*. Indianapolis, Indiana. June 2014. Available at https://peer.asee.org/22949 [Accessed 10 January 2020].

Spindler, R., 2019. Aligning Modelling Projects with Bloom's Taxonomy. *PRIMUS*, pp.1-17. http://dx.doi.org/10.1080/10511970.2019.1619208

Struyven, K., Dochy, F. and Janssens, S., 2005. Students' perceptions about evaluation and assessment in higher education: a review. *Assessment & Evaluation in Higher Education*, 30(4), pp.325-341.

Suskie, L., 2004. *Assessing Student Learning (A Common Sense Guide)*. San Francisco: Jossey-Bass.

Thorne, S., 2016. *Interpretive Description: Qualitative Research for Applied Practice*. New York: Routledge.

Zimmerman, B., 1989. A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, 81(3), pp.329–339. http://dx.doi.org/10.1037/0022-0663.81.3.329