Project work for a large class - Authentic assessment

Ayse Aysin Bilgin[1], Jaslene Huan Lin[2] and Nazli Ersahin[1]
[1]Macquarie University, Australia
[2]University of Sydney, Australia
ayse.bilgin@mq.edu.au

Improving graduate employability is increasingly essential due to evolving industry demands. Through experiencing the data investigative workflow comprising framing research questions, implementing data analysis and communicating results, students develop key employability skills such as organisational and time management skills. Large statistics classes pose challenges for fostering engagement and assessing practical skills authentically. This paper presents a scalable approach using authentic data for a project assessment that enables students to make decisions for their data investigative workflow. Examples of student work indicate that project-based assessment deepens understanding and strengthens connections between theory and professional practice. These experiences help build critical thinking and communication skills which are vital for today's job market. The paper concludes with implications for assessment design and acknowledges current limitations, offering a roadmap for educators aiming to bridge classroom learning and employability through authentic experiences.

INTRODUCTION

Authentic assessment in statistics involves tasks that mirror real-world problems, requiring students to pose or finetune a research question in the context of the problems and data at hand, understanding the relevant data before they apply their statistical knowledge and present their results (American Statistical Association, 2016). One way to implement authentic assessment in statistics is to design project-based assessments. Researchers over the last three decades have investigated ways of implementing authentic project-based assessments and their benefits to statistics education. While some researchers focused on improving one employability skill such critical thinking (Bates et al., 2025), others showed that implementation of project-based learning and assessments could enhance student engagement and motivation (Bilgin et al., 2015); enhance perceived relevance of statistical concepts, leading to improved learning outcomes (Neumann et al., 2013); improve students' statistical reasoning and their confidence in handling data-driven tasks (Awuah et al., 2021); enable students to develop positive attitude towards statistics and increase self-efficacy in conducting statistical analyses (Elder, 2023); provide opportunities to move away from traditional procedural lower order thinking assessments to skills based higher order thinking assessments (Chance, 1997); enhance students' employability skills and overall learning experience (Bilgin & Prvan, 2021).

The scholarly discussions on how statistics education could benefit from changes are not new. The scholarly publications on such discussions includes hands-on learning using projects (Bentley, 1994); experiential learning to enhance student engagement and understanding so that they become lifelong learners (Weldon, 2014); and utilising computers to deepen statistical learning and reasoning (Garfield & Ben-Zvi, 2008). Experiential learning in statistics has evolved over the decades, linked to the introduction of computer labs in the early 1980s, though access to digital resources remains limited in some institutions. Nevertheless, experiential learning, as defined by Kolb's (Kolb, 1984) cyclical process of experiencing, reflecting, thinking, and acting can be implemented without technology, as illustrated by Deming's "Monte Carlo experiment with the funnel" (p. 327), which demonstrates the costly effects of overadjustment due to misunderstanding common versus special cause variation.

The focus of this paper is experiential learning (Kolb, 1984) by using an authentic project-based assessment, the focus is not designing project-based learning. Experiential learning emphasises the cyclical process of concrete experience, reflective observation, and active experimentation, aligning well with the project assessment used in this study. Authentic assessment focuses on evaluating students through tasks that mirror real-world challenges, thereby promoting deeper engagement and meaningful application of statistical concepts. Integrating these frameworks fosters both cognitive and practical competencies, ensuring learning is reflective and transferable. We argue that even if classroom learning

In: Birk, L., Loth, G., Jotzo, L., Binder, K., & Frischemeier, D. (Eds.) (2026). *Statistics and Data Science Education in STEAM. Proceedings of the Satellite Conference of the International Association for Statistical Education (IASE), September/October 2025, Münster, Germany*.

is not designed to be project-based learning, project-based assessment would benefit student learning outcomes, if the assessment is carefully designed and implemented.

BACKGROUND

The Business School at Macquarie University decided to change the curriculum for all their units based on industry consultations in 2022. The first two authors were part of a team that revised the first year Business Statistics unit curriculum. The unit's learning outcomes have been updated to better reflect current industry needs and expectations. The revision placed greater emphasis on active learning and the development of practical technical skills, particularly Excel mastery, a widely used tool in the business sector. The unit title was also changed to *Business Statistics and Insights* to highlight the importance of engaging with data through hands-on statistical problem-solving to derive insights. Learning activities, a critical component in scaffolding student learning, have been redesigned to focus less on formula-based calculations and more on applying statistical techniques within Excel. In alignment, the assessments were revised with the aim of improving employability skills of students by encouraging them to apply their knowledge. The new assessments were pre-workshop and in-workshop quizzes (20%) to encourage weekly engagement and learning; a project report (20%) where learnings from earlier weeks were consolidated and applied to real data by week 8; a class test (20%) in computer laboratories by utilising randomised questions created using R-exam package (Bilgin & Lin, 2021) towards the end of the session and a final exam (40%) which used the same data set for all the questions and students were provided with the data set and a series of questions during the session. The project assessment, which is the focus of this paper, is presented in more detail in the next section.

THE PROJECT ASSESSMENT

The European Social Survey, conducted regularly to update available data for public use, was the source of the data set used for the project. The larger version of the data was used during lectures and tutorials, so the students were already familiar with the data set. The Excel file had two sheets: data (n = 25,534) and meta data for the variables. Two variables were created by the teaching team so that students could have numerical variables for their analysis. The variables were

- *cntry*: country - three countries: Spain, Netherlands and Bulgaria (nominal categorical)
- *nwsptot*: newspaper reading, total time on average weekday (ranges from 0: no time to 7: more than 3 hours)
- *netuse*: personal use of internet/e-mail/www (ranges from 1: never use to 7: everyday)
- *chldhm*: children living at home or not (binary)
- *nwsptot_N*: newspaper reading, total time on average weekday in hours (numerical, continuous)
- *netuse_N*: personal use of internet/e-mail/www hours in a week (numerical, continuous).

To help students establish a clear link between the assessment and how it facilitates the attainment of the unit's learning outcomes, the assessment started with a brief blurb to set the context. Students are advised to think from the perspective of an entrepreneur who is going to create a product for the market. Given the data set and the variables, the students were asked what product they would create, and what their business/research question would be, followed by a description of the variables in the data set. And then given instructions to choose two of the three countries, three of the five remaining variables (countries are already chosen) and one of the four generic questions. The generic research/business questions were: 1) *Is there an association/relationship between the two variables for the two countries?* 2) *Does the average xxx differ between two countries?* 3) *Is the proportion of people xxx equal to 30% in one of your chosen countries?* and 4) *Does the proportion of people's personal use of internet/e-mail/www or newspaper reading, total time on average weekday depend on them living with children or not in one of your countries?* We aimed to provide students with the autonomy to choose a research question that aligned with their interests, while also applying some constraints to ensure consistency and fairness in marking. Each research question included hints to specify what is expected of the students, such as for research question 1, the hint was "Use your chosen two countries and choose 2 of the variables for this question. You need to create your data explorations, graphics, tables, Excel explorations and if possible, statistical analysis to answer this research question.". This part of the assessment was used to set the limits for the student's analysis, with a requirement to undertake

preliminary data analysis by creating graphs and numerical summaries before answering the assessment questions which they will be marked on.

The assessment questions guided the students to structure their answers, by telling us which product they have chosen, which research question they were answering (note that research question needed to be modified and specific to their research question), data explorations, including how they dealt with missing or outlier observations, graphical explorations (only four needed to be included here, remaining could be placed in the appendix), identifying statistical measure of centre and variability suitable for the numerical variables and why, tables for summary statistics or count data, Excel explorations such as named ranges, slicers, a business decision and conclusion based on their analysis (e.g. answering the research question, ideally using hypothesis testing). For the Excel explorations, they needed to show screenshots in a 2-page Appendix 1 of what they had done. Appendix 2 was for letting us know if they used Generative Artificial Intelligence tools and what they found useful or not (one page only). Finally, Appendix 3 was for anything else they wanted to show us, such as extra graphics or analysis with no page limit for this appendix. We provided an empty Word template where the page length for the answers to assessment questions was set to be 4 A4 pages, with the 5th and 6th pages for Appendix 1, the 7th page for Appendix 2, and Appendix 3 starting on the 8th page.

One week after the assessment release, many students had questions seeking clarification. In response, we provided Assessment guidelines, which included alternative research questions and a detailed marking scheme, showing specific marks allocations, to support student understanding, as for most, it would be their first independent, student-driven, open assessment.

EXAMPLES OF STUDENT WORK

Students submitted their answers file as pdf and their Excel file with their analysis. The pdf file is marked. We only looked at Excel file, if there was any doubt about students' work. We prepared solutions for the markers for each possible question using different combinations of two countries using the full data set.

The products students considered are interesting. Some examples of the products are "NewsFlash is a subscription-based news app that provides readers with concise, summarized news articles directly to their smartphone.", "PlanPal: The Ultimate Organizational Kit for Kids, which offers a comprehensive suite of interactive tools including digital calendars, magnetic task boards, and vibrant stickers.", "a content creation course aimed at writers, marketers and journalists", "a news tabloid in the form of a website which presents the latest news in Soccer in all age competitions." and "launch a new electronic reading device which stores countless books and newspaper articles aimed at the European demographic". However, some students ignored the data set when considering products therefore follow up business questions could be considered using the variables in the data set were either superfluous or irrelevant, such products included "an automatic toilet cleaner".

The business/research questions posed by the students "Does the average newspaper reading time differ between Bulgaria and Spain?", "Does the average personal use of the internet, email and the world wide web differ between Spain and Bulgaria?" and "Is there an association/relationship between 'nwsptot' (newspaper reading time) and 'childhm' (children living at home or not) for Spain and The Netherlands?". Some questions were wrongly worded "Is there an association/relationship between the internet use in Spain and the Netherlands?" which should have been "is there a difference …" not association or they were too general "Is there an association/relationship between the two variables for the two countries?" it is not clear which two variables and which two countries are considered here.

Data explorations, graphical displays, tables, excel explorations were mostly done following what the students learnt in their lectures and tutorial. They applied their knowledge to an unknown data set for a question they created. They needed to consider which graphical displays to use, what goes on to x and y axes, they needed to make sure graphs and tables are informative by clearly labeling them. This part of the assessment was not so hard when they identified the graphical display, though the questions we received from students indicated that sometimes they were confused.

Students used the hypothesis test for comparison of two means between two countries (two independent samples t-test) to conclude that "the average time spent on newspaper reading per week does differ between Bulgaria and Spain"; for investigating whether there was an association between newspaper reading (ordinal categorical variable) and children living at home (binary) (Chi-square test

of independence) to conclude that "there is an association/relationship between 'nwsptot' (newspaper reading time) and 'childhm' (children living at home or not) for Spain and The Netherlands." since there was enough evidence to reject the null hypothesis of no-association. In the Chi-square example the wording could be improved by stating "there is evidence to suggest an association …" but considering these were the first-year students, such answers were not penalised, instead we gave them feedback to be more careful when they are explaining their results.

The business decision and conclusion were not always based on hypothesis testing, instead exploratory data analysis was used for the basis of their answers. Even so, sometimes the conclusion was very vague "it was observed that there is a potential association between internet usage and the number of children living at home in Spain and the Netherlands. The data suggests diverse patterns of internet consumption among households, influenced by family demographics." such answers are no better than using fortune telling since they do not include any details of what to expect such as "What kind of association was observed?" "What were the diverse patterns?".

There were 13 students out of 992 who achieved full marks for the project report, 78 did not submit a project, 7 received a zero mark (e.g. 5 of these were academic integrity breaches, two were genuine zeros). Three-quarters of students achieved a passing mark (n = 724). The average mark out of 20 was 12.5 marks with a standard deviation of 4 marks. The distributions of all assessments show that quizzes (max. 20 marks) and class test (max. 20 marks), which were created using R-exam package, were left skewed while the distribution of report marks (max 20 marks) and final exam marks (max. 40 marks) were more like normal distribution (Figure 1).

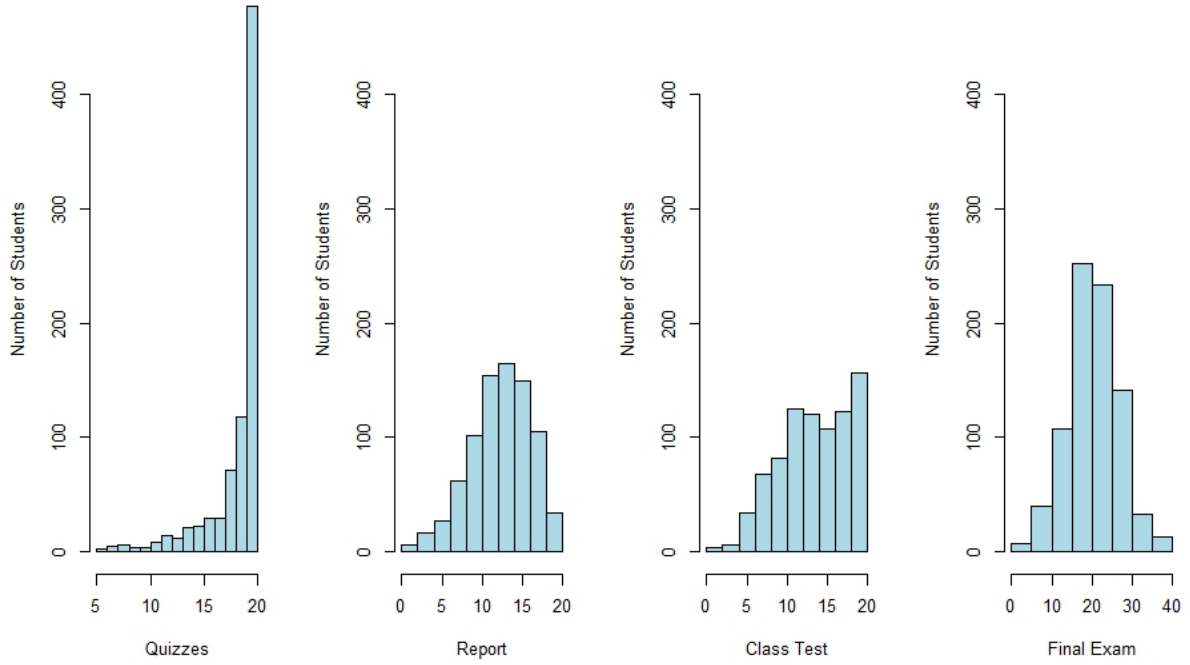

Figure 1. The distribution of four assessment tasks.

STUDENT FEEDBACK

Less than one-third of the students completed the official unit evaluation survey (n=261). More than half of the responders agreed or strongly agreed that the unit helped them to develop work-related professional knowledge and skills (62.5%), critical thinking skills (57.7%), problem solving skills (64.7%) and close to half of the responders agreed or strongly agreed that the unit helped them to develop communication skills (45%).

There was good news such as "This unit assisted me with my job. Idk *(I don't know)* if people will like it or not though" and some constructive student feedback such as "Please provide an actual example of the report instead of a template.". However, despite all our efforts to clearly explain what is needed, additional guidelines, holding extra help sessions and a one-week extension to submission

deadline, written students' comments were mostly heart breaking. Such negative comments included "I prefer quiz than report", "Assignment (*project report*) was very confusing and not explained.", "This is a good and interesting unit, however it was difficult to understand some of the content. I found the assessment task (*project report*) quite confusing and a bit difficult to complete." and "The assignment in the report was very unreasonable, because there were a lot of data".

We also had 3 student representatives who gathered anonymous feedback. Based on their report, 32 out of 45 students (71.1%) who they surveyed strongly agreed that they would prefer "lectures to have a greater emphasis on practical skills …". Even though we tried very hard to make the instructions, guidelines and hints on project report assessment files to be as clear as possible, nearly half of the responders of student representatives' survey disagreed or strongly disagreed that "the assignment's rubric was clear, and they knew what is expected of them" (46%), while one-third agreed or strongly agreed (29%). It is interesting to note that 44% of the responders agreed or strongly agreed that "they knew where to find resources that will help in their assessment" while 34% disagreed or strongly disagreed.

Some students wanted to have an example report instead of a template, they argued that they would know better what their report is meant to look like if they can see a report. The first author used anonymised previous student reports as example reports for a post-graduate unit which seemed to reduce student anxiety. Experience showed that one or two example reports would be more than sufficient because the students do not want to invest time into reading through previous students' reports, but they want to see at least one to give them a sense that they are on the right track. Even though it seems the example reports could limit creativity, for the first-year large classes, it is recommended that example reports are made available to the students.

DISCUSSION AND CONCLUSION

This paper shows the results of implementing a data analysis project using a large real-world dataset with diverse scope where students were encouraged to think creatively in a large first-year service unit. We guided students with "structured" open-ended statistical questions. Students were required to pose their research questions, choose their variables, explore the data using descriptive statistics and graphics, then answer their chosen research questions via hypothesis tests which are aligned with GAISE college report (American Statistical Association, 2016).

While this study focused on a single unit within one institution, student assessment examples highlight how students developed employability skills such as posing a business question, using available data to find an answer for the chosen question, communicating their results. Student feedback demonstrates the need to be more specific about what is expected of the students when they are given a project assessment to promote deeper learning, so that the gap between theoretical knowledge and professional practice can be closed. There is no doubt that authentic assessment supports the development of critical thinking and communication skills—vital for today's job market but how it is embedded in a unit could make it great or cause a lot of tension between students and teachers, as a result students could become very unsatisfied. Keeping in mind that "…recruiters often weigh assessments of real-time problem-solving or proof of project-based achievements more heavily than a traditional diploma or degree." (Schreiner, 2025), we should not be discouraged and move away from such useful assessments, instead we should persevere and help students to become better learners.

In the era of widespread Generative Artificial Intelligence use, it is inevitable that higher education learning and assessments will change, as it did when calculators and computers found their place in learning and assessments. At the end of the day, the aim of higher education institutions is "... building the competence of their students to think critically, to be creative, to collaborate and communicate effectively – the well-versed core 21st Century skills." (Schreiner, 2025).

Our explorations of different assessments, online quizzes completed outside classroom, online tests completed in class, invigilated final exam and project report assessment showed that the performance of students in project assessments were similar to the invigilated final exam (Figure 1). We argue that GenAI is not yet mature enough to provide answers to such authentic assessments with large data sets, and they can be used instead of invigilated exams. However, we also noted that students are disgruntled with such assessments since they seem to want, step by step, recipe book style assessments where at each step they are told what to do which is not fit for what is expected of higher education

graduates. Education should start with helping students to transition from structured high school education learning to higher education individual learning where responsibility of learning lies with the student, not with the teacher.

To evaluate the pedagogical impact of the project assessment, we reviewed student submissions based on their ability to formulate a clear research question, select appropriate variables, apply relevant statistical methods, and communicate their findings effectively, including the use of appropriate graphical displays. We expected the report to be coherent overall, with different sections well connected and logically structured. Students demonstrated learning through their capacity to explore data independently, apply their statistical knowledge, and draw evidence-based conclusions. While the quality of work varied, many students showed thoughtful engagement and creativity in their approach. These outcomes suggest that project-based assessments can support the development of higher-order thinking and transferable skills. From a teaching and learning perspective, this reinforces the importance of authentic tasks in helping students connect theory with practice, particularly when expectations are clearly communicated and learning is scaffolded throughout the semester.

From a pedagogical standpoint, the integration of experiential and authentic assessment strategies fostered a learning environment that encouraged autonomy, critical thinking, and real-world relevance. However, despite the efforts to foster student engagement and simulate authentic workplace experiences through open-ended, project-based data analysis tasks, many students did not fully perceive the value of the assessment. As reflected in student comments such as *"*... the mid-session assessment *(project report)* was unnecessarily difficulty and was not very clear nor concise as to the expectations*... "*, the purpose and relevance of the task were not always clear to them. Students need to understand why such assessments are used and how they can benefit from them to become employable otherwise they resent the assessment and feel under pressure then become anxious.

This highlights several important implications for future designs of project-based assessment for first-year students. First, the purpose and relevance of the assessment must be explicitly communicated to students. Gaining their buy-in early is critical for meaningful engagement. Second, providing comprehensive assessment guidelines, including sample questions and clear marking criteria, can proactively address common student concerns and reduce uncertainty. Finally, to better support students in building toward the assessment, it is worth integrating bite-sized scaffolded activities into lectures and tutorials. These activities can help students gradually develop the component skills needed, which they can then integrate in the final project. In summary, implementing project-based assessments in large first-year units is a daring but necessary step towards cultivating statistically literate and work-ready graduates equipped to thrive in a data-driven environment.

ACKNOWLEDGEMENTS

We thank Dr. Prashan Karunaratne for his enthusiastic contributions to curriculum revision and his passionate teaching. We are also grateful to Ms. Nikhara Inderlal for her tireless support in responding to student queries and assisting with the coordination of our large team of tutors.

REFERENCES

American Statistical Association. (2016). *Guidelines for assessment and instruction in statistics education (GAISE) college report*. https://www.amstat.org/docs/default-source/amstat-documents/gaisecollege_full.pdf

Awuah, R., Gallagher, K. M., & Dierker, L. C. (2021). Taking Project-Based Statistics Abroad: Learning Experiences and Outcomes of a Project-Based Statistics Course in West Africa. *Statistics Education Research Journal*, *19*(3). https://doi.org/10.52041/serj.v19i3.54

Bates, J., Sherri, C., Mark, F., & Wang, X. (2025). Cultivating Critical Thinking Skills: A Pedagogical Study in a Business Statistics Course. *Journal of Statistics and Data Science Education*, *33*(2), 166–176. https://doi.org/10.1080/26939169.2024.2394534

Bentley, D. (1994). Hands-on and Project Based Teaching: Sessions 4A, 4B, 4C. In The National Organizing Committee of the Fourth International Conference on Teaching Statistics (Eds.), *The Fourth International Conference on Teaching Statistics*. Rabat, Morocco.

Bilgin, A. A., & Lin, H. (2021). Designing Assessment Tasks To Prevent Cheating In A Large First-Year Statistics Unit. In R. Helenius & E. Flack (Eds.), *Proceedings of the Satellite Conference of*

the International Association for Statistical Education (IASE). August/September 2021. https://doi.org/10.52041/iase.errob

Bilgin, A. A., & Prvan, T. (2021). Different approaches to group project work in statistics classes. In A. A. Bilgin & S. Budgett (Eds.), *OZCOTS 10, Proceedings of the 10th Australian Conference on Teaching Statistics*, July 2021 [Virtual].

Bilgin, A. A., Newbery, G., & Petocz, P. (2015). Engaging and motivating students with authentic statistical projects in a capstone unit. In M. A. Sorto (Ed.), *Advances in Statistics Education: Developments, Experiences, and Assessments. Proceedings of the Satellite Conference of the International Association for Statistics Education (IASE)*, July 2015, Rio de Janeiro, Brazil.

Chance, B. L. (1997). Experiences with authentic assessment techniques in an introductory statistics course. *Journal of Statistics Education*, *5*(3). https://doi.org/10.1080/10691898.1997.11910596

Elder, A. C. (2023). Statistics Attitudes After Using Guided Project-Based Learning as Andragogical Strategy in a Graduate Statistics Course. *Statistics Education Research Journal*, *22*(3). https://doi.org/10.52041/serj.v22i3.436

Garfield, J., & Ben-Zvi, D. (2008). Chapter 5: Using Technology to Improve Student Learning of Statistics. In J. Garfield & D. Ben-Zvi (Eds.), *Developing students' statistical reasoning: Connecting research and teaching practice* (pp. 91–114). Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-8383-9_5

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development* (Vol. 1). Prentice-Hall.

Neumann, D. L., Hood, M., & Neumann, M. 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). https://doi.org/10.52041/serj.v12i2.304

Schreiner, G. (2025, May 12). AI prompts us to rethink the very foundations of higher education systems. *Daily Maverick*. https://www.dailymaverick.co.za/opinionista/2025-05-12-ai-prompts-us-to-rethink-the-very-foundations-of-higher-education-systems/

Weldon, K. L. (2014). Experience Early, Logic Later. In H. MacGillivray, B. Phillips, & M. A. Martin (Eds.), *Topics from Australian Conferences on Teaching Statistics* (pp. 25–42). https://doi.org/10.1007/978-1-4939-0603-1_3