# TECHNOLOGIES FOR ENHANCING PROJECT ASSESSMENT IN LARGE CLASSES

Michael Bulmer
School of Mathematics and Physics, University of Queensland, Australia
m.bulmer@uq.edu.au

*The use of technology in statistics assessment is widespread. These uses include assessment tasks that are moderated by technology, such as formative or summative online quizzes, as well as the more fundamental empowerment of students to be able to tackle realistic data sets and more sophisticated modelling in their assessment tasks through technology. In this paper we identify and survey a third role of technology, supporting project assessment in large classes, and give two key examples of this. The first will be the use of virtual environments to engage students with statistics in context, including issues in experimental design and measurement. The second will look at using technology to enable a one-day statistics conference where each student in a class of eight hundred can give a ten-minute oral presentation on the use of statistics in scientific research.*

## INTRODUCTION

Project work has many purposes in a statistics course. In a famous quote, Cobb and Moore (1997) suggest that “statistics requires a different kind of thinking [to mathematics], because data are not just numbers, they are numbers with a context.” Projects can provide that context for students, especially when the projects involve experimental or survey work whereby students are generating their own data. MacGillivray (1998) and Mackisack (1994) give an overview of the other benefits of experimental work, such as an appreciation of the practical issues involved in carrying out experiments and collecting data, an outcome encouraged by Higgins (1999).

In our setting, teaching introductory statistics to undergraduate science students, we further believe that project work is important for inducting students into scientific practice and providing first-hand experience in the role that data and statistical reasoning play in science. This goes beyond simply providing a context for the numbers and necessitates authentic activities that engage students in various modes of scientific inquiry.

The challenge then is how to facilitate such authentic experiences in large classes, particularly when the statistical learning embedded in these experiences is to be assessed.

## ISLAND POPULATION

For many years we have had students carry out real experiments and surveys as a means of generating data for analysis. Students designed and carried out the experiment, collected the data and conducted appropriate statistical analysis, and presented the results of their research in the form of a scientific paper, giving them a fairly realistic experience of statistics in the scientific method. However, we quickly ran into ethical issues in the kinds of studies that students wanted to do. Overall this has been a positive development, with an ethics component now added to the course, but it has severely limited the kinds of experiments students could undertake. A common outcome was for students to dissolve headache tablets in hot and cold water and compare the dissolving times–a useful enough “context” for statistical analysis but hardly an authentic scientific experience.

Our solution to these issues has been to develop an online environment where students can conduct studies involving virtual human subjects. One of the immediate aims of this was to free students from real ethical concerns (although they could certainly still engage in the virtual ethics) but we have found many other benefits of using this environment for assessing student learning.

### *Epidemiology*

The online environment–the Island–gives a map with 38 villages. Each of these villages contains households of living Islanders as well as a cemetery where users can view details of Islanders who have died in the village. Another motivation for this technology was the need for a virtual environment where students could collect data for questions in epidemiology. While it is easy to generate some example data for a randomized clinical trial, for example, our belief is that thinking about issues in epidemiology requires access to a population that can be studied more

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php 

deeply. For example, to determine whether diseases have a strong genetic component you might want to consider the ancestors of particular individuals. The Island involves three intertwined simulations running at different timescales with the longest timescale giving a history or births and deaths on the Island for around 240 years. The combined population (living and dead) that students can use in their studies is 14,771 Islanders.

*Islander Images*

A key visual component of the interface for a particular islander was including a computer-generated image of them. We used an improved version of the system developed by Bulmer and Engstrom (2005) to generate full-body and facial images for each of islanders. Sample images of two of the islanders are shown in Figure 1.

In addition to helping students identify their study with "real" people and visually engage them with the interface, the images also served to clearly show inherited features by comparing individuals with their parents, for example. In particular, facial close-ups were included primarily so students could see eye colour clearly since it was sometimes hard to see from the full-body images.

Figure 1. Sample full-body images – Ian Lopez and Summer Quinn

*Reality versus Fantasy*

There is a fundamental tension in the design the Island. We want the simulation to be realistic since we believe that will help students engage with the virtual environment. However there is a point at which realism becomes counterproductive towards our aims of engaging students in the role of statistical reasoning in scientific inquiry. For example, suppose we included a cause of death called *Lung Cancer* and made it so that islanders with higher smoking levels were more likely to die from this disease. Students could collect data on smoking history and cause of death and look for this relationship but if they found an association it would not be surprising to them. They will not have discovered anything new by conducting their study.

Instead of using real names for diseases we have thus tried to use poetic names wherever possible. These include *Summer's Pain* (named for Summer Quinn, the first person on the Island to die of that condition, and also shown in Figure 1), *Diego's Cough*, *Ruin* and *Jungle Sickness*. One of these four is indeed modeled on lung cancer, including the association with smoking history, but now it is a more open question for students to explore. For example, what data do you need to collect to distinguish between these conditions and how can you convince somebody that you have identified 'lung cancer' on the Island?

*Collaborative Design*

A key feature of the Island has been the involvement of students in its creation. The base simulation described above gave a rich population to explore but one that was essentially static.

The first semester of use with science students involved an assessment task where each student had to prepare a research proposal with the islanders as their subjects. The main part of this proposal was specifying the kinds of treatments and measurements that the study would involve.

There were very few constraints on the kinds of proposals allowed since the overarching aim of the project work was to accommodate a wide diversity of student interests. For each student proposal that required an addition to the Island we began by searching for existing research on the topic. This gave plausible ranges for response variables as well as suggesting relationships that might be included in the simulation.

The other two simulations on the Island then run at shorter timescales: one daily and one ticking every thirty seconds. These simulations incorporate a variety of models to meet the needs of the students. One of the earliest components added was a simple model of glucose and insulin dynamics using a pair of differential equations (Yipintsoi et al., 1973; Fisher & Teo, 1989). This enabled studies where students wanted to look at the effect of foods with various glycemic indices on blood glucose as well as other proposals that wanted to compare injections of synthetic and natural insulin. An islander could be made diabetic by setting one model parameter close to 0, allowing a range of studies that wanted to compare diabetics and non-diabetics in various ways.

As before there is tension between reality and fantasy here. Making the simulated processes perfectly match reality would be technically difficult and, as with the smoking and lung cancer example, may not actually be desirable. We felt it was important to keep students on their statistical toes by omitting some associations that they might expect to find while adding some other associations that would surprise them, though we did keep this at a low level. A better alternative for the long term is to add tasks that are somehow native to the Island. For example, we added ‘dalpa leaves’, native to the Island, and allowed islanders to “chew lime-soaked dalpa leaves for ten minutes”. This was initially added as a control for chewing lime-soaked coca leaves (a treatment requested by a student) but dalpa leaves were given their own effects that students can study independently in the future.

*Surveys*

In addition to tasks that mostly behaved like measurements, the students could also design a survey for their Islanders to complete. Questions that students were interested in asking were added to the system, along with a range of standard survey questions that we typically ask the students themselves (such as age, height, weight, eye colour, which superpower they would most like to have and how attractive do they think they are to members of the opposite sex). The Islanders would take longer to complete longer surveys, discouraging students from just asking all the questions. Some islanders were predisposed to lie on the surveys, particular for questions related to age and weight.

Eighteen of the survey questions come from the shortened version of the Profile of Mood States (Shacham, 1983). Students could use these to measure emotional and attitudinal responses to the various stimuli.

*Student Feedback*

The student feedback was overwhelmingly positive. Many students commented on the obvious aim of the islanders in that “they were interesting and a great way to find results of experiments. It made the experimentation process nice and easy to conduct”. There were also comments indicating that students were connecting with their virtual subjects: “I liked how we were able to see their whole history on their profiles; it was interesting seeing some of their troubled past.”

Almost all the negative feedback was related to sleep. Each night the islanders went to sleep, using a hypnogram simulation based on a Markov chain model developed by Kemp and Kamphuisen (1986). Many students were interested in sleep deprivation and so were able to prevent the islanders sleeping in some way (as long as they stayed up with the islanders to keep them awake). However, most students simply wanted to do other experiments and were frustrated that their subjects were asleep. We again feel this important, another limitation faced by experimenters in practice. We did concede some ground though, changing the simulation so that the islanders went to sleep an hour later than originally designed.

*Publication Bias*

One of the most interesting pieces of feedback in the semester came from the following student email: “I am doing my experimental project and I’m trying to find a change in anything basically after my islanders take marijuana. First I tried to give them cannabis tea and waited ten minutes and measured pulse, IQ and gave them my survey but there weren't any significant changes between before and after. So today I tried again and gave them a reefer instead, waited 10 minutes and again no changes. Is ten minutes too long or not long enough? I’m using a variety of age and sex so it’s not that they are too young or old. Does it matter that the changes aren’t very good? ... Will I lose marks for this?”

Apart from illustrating the student’s engagement with the task, this was a perfect example of the pressures leading to publication bias in scientific practice and we’ve made use of this quotation (with the student’s permission) as an example in lectures. It is in this way that the Island helps to support assessment of statistical thinking, placing students in complex settings where they have to make serious use of statistics to find answers.

STUDENT CONFERENCE

As mentioned, we can view the aims of our teaching as engaging students in scientific practice. With the virtual environment above the focus was on the practices of hypothesis formation, experimental design, data collection and statistical analysis, and written communication of a scientific study. In a complementary project task we ask students to find a published scientific paper in a peer-reviewed journal. This paper can be in any area of interest to them, again catering for the diversity of interests in the large class.

*Conference Structure*

It is common for students to give oral presentations in small group tutorials but there are practical issues with doing this. For example, we typically have small group classes of 20-25 students and so to allow every student to give a ten-minute presentation would require 4 or 5 hours of the class time. In some cases this can be spread across the semester to be a valuable learning experience within the small group program. However, our aim with this task was for students to use the presentation to improve their holistic understanding of statistical reasoning and so we wanted it to happen later in the course. Given other pressures for the small group classes, such as helping with analysis for the experimental project work discussed earlier or preparing for the final examination, we did not feel this was appropriate.

Another option for making oral presentations more manageable is to have the presentations done in groups. This simply means that individuals have smaller roles in the actual presentations. Moreover, while being able to work in a team is an important graduate outcome, in scientific practice it is still most common for seminars and conference presentations to be given by an individual who can provide an overview of the ideas, whether they come from their personal work or from the work of a research team. As with the experimental project, our ongoing motivation for project work in large classes is to provide an authentic engagement with scientific practice. Thus we have opted for running oral presentations in the form of a one-day conference where all students would give individual talks.

*Conference Organization*

In 2009 we ran the conference twice, once in each semester. Based on the success of previous semesters, in Semester 1 we invited students doing similar introductory statistics courses in two other programs (pharmacy and physiotherapy) to participate in the conference, rebranding it as the institution’s “Undergraduate Statistics Conference”. The combined enrollment in the three courses was 770 students.

Each student was giving a ten-minute timeslot for their presentation. This consisted of five minutes for the presentation itself, three minutes for questions and discussion, and two minutes for changing over to the next speaker. This required a total of 7700 minutes of timeslots, around 128 hours.

To achieve this we broke the day into four blocks, each of two hours duration. Within a particular venue we scheduled 10 students to give presentations during each block, leaving 20 minutes spare in the block to allow for any technical problems and for changing over to the next group. We needed 77 of these groups of 10 students, requiring up to 20 parallel sessions during each block.

The only way we could access this number of venues (each requiring a computer and data projector) for a whole day was to run the conference on a Saturday. An online registration system was used to allow students to choose which of the four blocks they wanted to attend (since many had work or sporting commitments). Registration opened around a month prior to the day of the conference. Importantly, as discussed below, students were then randomly assigned to the various parallel sessions in each block.

*Peer Assessment and Feedback*

The key innovative use of technology comes in managing the assessment of the student presentations. We investigated a variety of assessment options and available technologies before settling on one combination that has proved reliable and effective. While online technologies seem the default solution to many problems, here our approach is based on paper.

In each session we have ten students giving presentations, each in turn to the remaining audience of nine students and the tutor chair. As they are the audience, we have the students peer assess the presentations they watch. For each presentation and each student reviewing the presentation we generate a sheet of paper on which the student records their assessment. This has the name of the student reviewer on the top along with a four-character code to identify the presenter (so the presenter can remain anonymous if they wish). A section of multiple-choice bubbles allows the reviewer to quickly rate the presentation against standards for five criteria. There are then two boxes for hand-written feedback. The first asks the reviewer to justify their assessment based on the criteria. Many students achieve a high standard on these criteria and so the second box asks the reviewer to suggest one or more ways in which the presentation could be improved (even if it was “perfect”).

At the bottom of each sheet is a barcode to identify the combination of presenter and reviewer. Once the conference is finished we scan the 7700 sheets of peer assessments along with the 770 sheets of tutor assessments. This takes several hours but once it is done all the students can directly access their feedback online. Thus, typically within 48 hours of the conference being held, each can access a compilation of written feedback from nine peers plus the tutor chairing the session. This is much more feedback than an individual tutor could provide in a large class setting and is very timely. The evidence from the conference data is that peers on average mark the same as tutors.

This feedback model is particularly effective because of the large numbers of students involved. The random allocation of students to sessions means that presenters will usually not know the students in their audience and so the feedback they receive is anonymous. Peer assessment in small classes, where students do know their audience and may recognize handwriting, adds extra social constraints to the feedback given.

## CONCLUSIONS

The student conference shows a project that does not in itself rely on technology (ignoring the fact that all students used PowerPoint for their presentations!) but where technology is vital in the assessment of the project outcomes. It also illustrates a use of assessment technology that is not just online–the success of the assessment approach has come through the integration of the paper assessment sheets, quick and reliable for use on the day, with the online mechanism for delivering the rich anonymous feedback from peers within a short timeframe.

Similarly, the Island aims to do nothing more than provide a flexible environment for carrying out virtual experiments. For example, it does not provide tools for managing or analyzing data, any more than a subject in a real experiment would. A project that uses the virtual environment does not need to be based on technology–it could be a traditional pencil-and-paper report with hand calculations. However, the technology of the Island aids in the assessment of the statistical skills by opening up students to the complexities of real scientific studies and making the required statistical thinking less routine.

ACKNOWLEDGEMENTS

This work would not have been possible without the input from students in suggesting the tasks for the Island. We are grateful for our research assistant, Kim Haladyn, who was responsible for researching existing literature on the physiological effects of these tasks.

REFERENCES

Bulmer, M., & Engstrom, C. (2005). Virtual Humans for Teaching and Learning Statistics. In I. MacColl (Ed.), *Proceedings of the Apple University Consortium Academic and Developers Conference*, 3.1–3.8. Online: www.auc.edu.au/conf/conf05/pdf/AUC_Conf_2005_Proceedings.pdf.

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

Fisher, M. E., & Teo, K. L. (1989). Optimal Insulin Infusion Resulting from a Mathematical Model of Blood Glucose Dynamics. *IEEE Transactions on Biomedical Engineering*, *36*, 479–486.

Higgins, J. (1999). Nonmathematical statistics: a new direction for the undergraduate discipline. *The American Statistician*, *53*(1), 1–6.

Kemp, B., & Kamphuisen, H. A. C. (1986). Simulation of Human Hypnograms Using a Markov Chain Model. *Sleep*, *9*, 405–414.

MacGillivray, H. (1998). Developing and Synthesizing Statistical Skills for Real Situations through Student Projects. In *Proceedings of the 5th International Conference on Teaching Statistics*, 1149–1155. International Statistical Institute.

Mackisack, M. (1994). What is the Use of Experiments Conducted by Statistics Students? *Journal of Statistics Education*, *2*(1). Online: www.amstat.org/publications/jse/v2n1/mackisack.html.

Shacham, S. (1983). A Shortened Version of the Profile of Mood States. *Journal of Personality Assessment*, *47*, 305–306.

Yipintsoi, T., Gatewood, L. C., Ackerman, E., Spivak, P. L., Molinar, G. D., Rosevear, J. W., & Service, F. J. (1973). Mathematical Analysis of Blood Glucose and Plasma Insulin Responses to Insulin Infusion in Healthy and Diabetic Subjects. *Computers in Biology and Medicine*, *3*, 71–78.