SIMULATING THE DATA INVESTIGATIVE CYCLE IN LESS THAN TWO HOURS: USING A VIRTUAL HUMAN POPULATION, CLOUD COLLABORATION AND A STATISTICAL PACKAGE TO ENGAGE STUDENTS IN A QUANTITATIVE RESEARCH METHODS COURSE

BAGLIN, James[1], REECE, John[2], BULMER, Michael[3] & DI BENEDETTO, Mirella[2]

[1]School of Mathematical and Geospatial Sciences, RMIT University, Melbourne, Australia
[2]Discipline of Psychology, School of Health Sciences, RMIT University, Melbourne, Australia
[3]School of Mathematics and Physics, University of Queensland, Brisbane, Australia

Contact email: james.baglin@rmit.edu.au

Abstract

Providing practical and relevant experience with the data investigative process is known to help engage students in quantitative research methods courses; however, limited student experience, time constraints, and ethical concerns present serious challenges to this goal. Recent advances in technology may provide innovative solutions. This paper reports on the development and implementation of a series of two-hour weekly computer laboratories that simulated the data investigative cycle in a research methods course for psychology students. The technologies included an online virtual human world, known as the Island, cloud collaboration technology and a statistical package. Each laboratory session aimed to engage students in all stages of the data investigative cycle, specifically: stating a problem, planning a study, gathering and analysing data and drawing conclusions. Student feedback from the sessions was overwhelming positive, which highlights the potential of combining innovative technologies to actively engage students in learning research methods and statistics.

INTRODUCTION

Understanding research methods and statistical analysis can assist students in making informed decisions in their everyday lives (Zablotsky, 2001) and can provide them with the necessary skills required to conduct their own research (Ball & Pelco, 2006). Regardless of these practical outcomes, students commonly perceive research methods courses to be difficult and uninteresting (Bridges, Pershing, Gillmore, & Bates, 1998; Gladys, Nicholas, & Crispen, 2012). Students often fail to see its relevance, and those with poor mathematical ability may exhibit strong anxiety towards the statistical components (Briggs, Brown, Gardner, & Davidson, 2009). These undesirable perceptions have been found to correlate negatively with a student's research methods course performance (Meldrum & Stults, 2012). Consequently, research methods instructors have long been interested in finding effective methods for better engaging students. Active learning strategies have been used successfully for this purpose.

Many research methods instructors have reported benefits of active learning approaches (e.g., Ball & Pelco, 2006; Barraket, 2005; Benson & Blackman, 2003; Crull & Collins, 2004). Active learning is based on the idea that a learner should be an active participant in the learning process (Bell & Kozlowski, 2008); however, actively engaging students in applied human research methods courses in areas such as psychology and social science, can be a challenge due to time, logistical and ethical constraints associated with human participants; therefore, students are often restricted to the extent in which they can be actively engaged in these courses. Self-directed project-based learning is an example of one successful active learning method for providing practical research experience; however, undergraduate students often lack the necessary experience to confidently initiate such projects. An intermediate step prior to self-directed learning would be beneficial for building fundamental research skills.

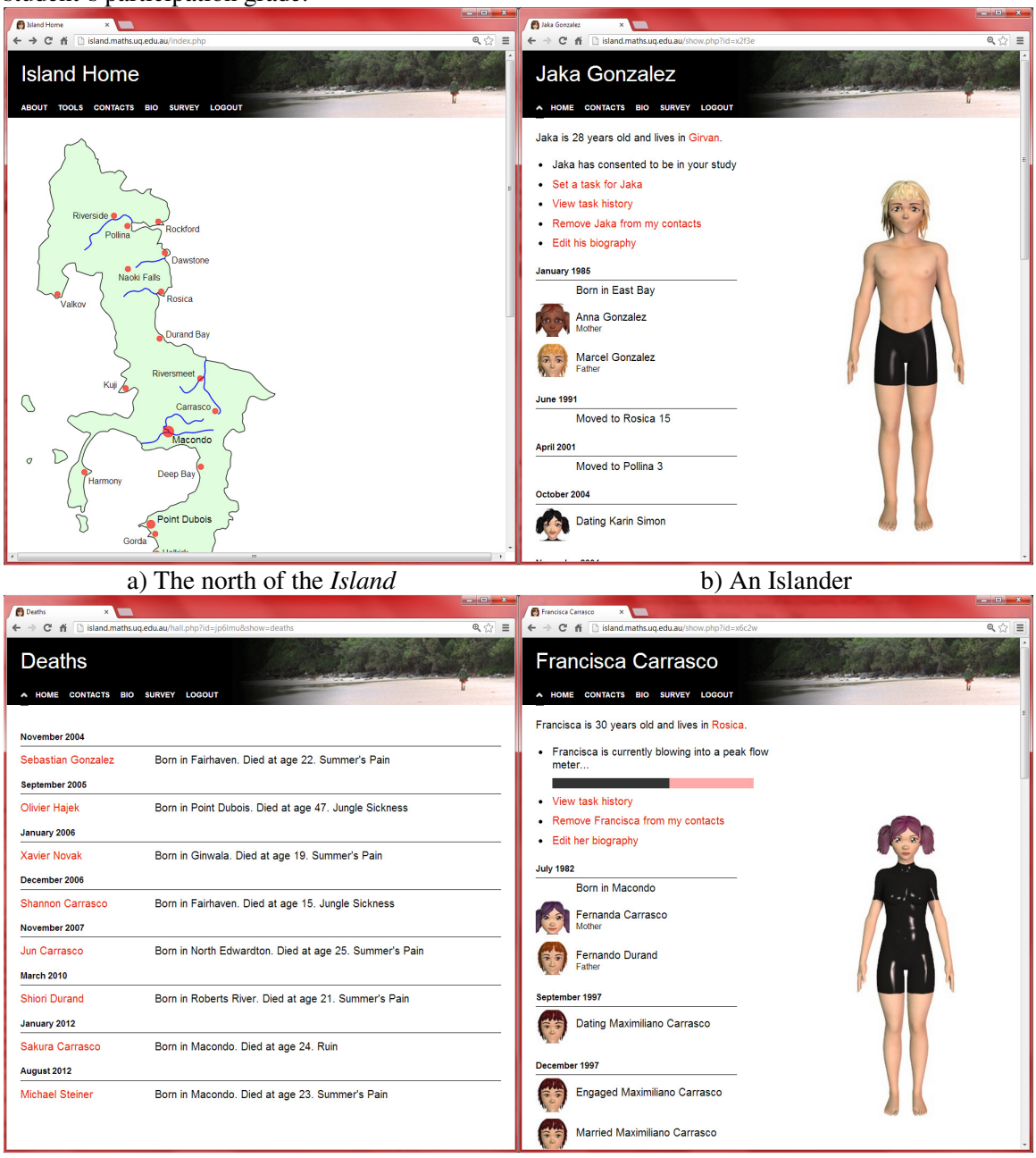
This paper reports on the development and implementation of a series of specialised computer laboratory sessions for a research methods course in psychology. These sessions aimed to simulate the research cycle within two hours using the innovative technologies previously mentioned. These sessions were developed to improve students' active participation during computer laboratory sessions, to build fundamental skills and experience required before independent research. While research methods and statistical analysis are topics in their own right, these sessions assumed that treating them together would enhance students' interest and understanding by providing proper context to both topics (Saville, 2008). The sessions were designed around the Problem, Plan, Data, Analysis and Conclusions (PPDAC) model of the data investigative cycle proposed by MacKay and Oldford (1999, as cited in Wild & Pfannkuch, 1999). This model guided the design of each computer laboratory session by linking research design and analysis into one seamless process. Three key technologies were used during these sessions. The virtual world, known as the *Island*, was used to propose relevant research problems and to plan and simulate data collection. Cloud collaboration technology, in the form of *Google* spreadsheets, was used for classes of students to gather data simultaneously into a central database. The statistical package, *SPSS,* was then used to analyse the data and reach conclusions about the research question being posed. Before the development and implementation of the sessions is described, a brief summary of the *Island* is presented.

The *Island*. The *Island* (http://island.maths.uq.edu.au) is a free, online, virtual human population that can be used for the purpose of simulating scientific data collection (Bulmer & Haladyn, 2011). Students navigate the *Island* by clicking between the 39 towns that are home to approximately 9000 "Islanders" (see
Figure **1** a.). Islanders are born, die, contract disease, date, marry and relocate in real time. Each Islander has a unique personal history and genetic make-up (see
Figure **1** b.). Islanders can be recruited for the purpose of conducting a wide range of scientific studies including surveys, observational studies, case-control studies, correlational designs and experiments. There are two major methods of data collection using the *Island*: archival and tasks. Archival information available on the *Island* includes births, deaths marriages, demographics (age, gender and residency), medical records (smoking history and past or current diseases) and family and relationship history (see
Figure **1** c.). Tasks are activities that *Islanders* will complete provided they consent to participate in a student's study (see
Figure 1 d.). There are currently in excess of 200 tasks. Examples of the many categories available include survey items (e.g., "How worthless do you feel right now?"), blood tests (e.g., sodium, type, and white blood cell count), physiological measures (e.g., respiratory rate, height, and weight), injections (e.g., saline, insulin and methamphetamine), mental tasks (e.g., mental arithmetic, reading comprehension, and grooved pegboard test) and exercise (e.g., shuttle run, arm strength and jogging). Statistical models control many of the effects, interactions and associations between tasks and archival information. The *Island* also provides a realistic research experience as Islanders are known to fall asleep, refuse consent, drop out during studies, get sick and lie.

THE COURSE

A total of 38 students (35 undergraduates and 3 postgraduates) enrolled in the psychology research methods course in the semester that the *Island*-based laboratory sessions were developed and implemented. The duration of the course was 12 weeks. Major assessment included an exam (40%), a methodology and data analysis assignment (30%), a lab report (20%) and tutorial participation (10%). The overall learning objective of the course was to introduce students to the major research methodologies and data analysis procedures used in applied psychological research. An overview of the topics included in this course is shown in Table 1. Weekly course contact hours included a two-hour lecture and two-hour tutorial session that took place in a computer laboratory. Computer laboratory sessions began in the fourth week of the course after students completed two standard classroom-based tutorial sessions. There were a total of nine computer laboratory sessions across the semester (see Table 1). Students were split into two

smaller classes for tutorials and laboratories, each supervised by a tutor. Tutors were trained to use the technology required for the *Island*-based laboratories for one hour prior to the first computer laboratory session. Attendance and participation in each session contributed to a student's participation grade.



a) The north of the *Island* b) An Islander

c) Archival information d) Tasks

Figure 1. The *Island* Interface

ISLAND-BASED COMPUTER LABORATORY SESSIONS

Island-based computer laboratory sessions were designed to provide students with a simulated experience of the PPDAC cycle within two hours by utilising the *Island*, cloud collaboration technology and a statistical package. Seven *Island*-based laboratory research topics were developed and implemented across nine computer laboratory sessions (see Table 1). The *Psychological Health Survey* topic was spread across three weeks. Session topics were lagged one-week behind lecture content. The first session, *Pilot Study*, was a basic introduction to using

the different technologies for each stage of the PPDAC cycle. In the following sections, the general design and implementation for each laboratory session is outlined in terms of the stages of the PPDAC cycle of empirical enquiry (see Figure 2 for an overview).

Table 1. Course Schedule and *Island*-based Computer Laboratory Sessions

Week	Course Topic	*Island*-based Session[a]	Brief Description
1	Introduction	N/A	N/A
2	Quantitative Research	Tutorial I[b]	N/A
3	Research Design and Hypothesis Testing	Tutorial II[b]	N/A
4	Introduction to *SPSS*	Pilot Study	Introduction to using the *Island*, *Google* spreadsheets and *SPSS*
5	Exploratory Data Analysis I	Psychological Health Survey I	Gather survey data from a random sample of Islanders
6	Exploratory Data Analysis II	Psychological Health Survey II	Screen and clean survey data in *SPSS*
7	Exploratory Data Analysis III	Psychological Health Survey III	Perform exploratory data analysis in *SPSS*
8	Comparing means	Gender and Self-estimated Intelligence	Quasi-experimental study demonstrating the use of a two-sample *t*-test
9	Correlation	Cannabis Use and Reaction Time	Repeated measures experiment demonstrating the use of a paired-samples *t*-test.
10	Categorical Associations	Correlates of Self-rated Attractiveness	Correlational study demonstrating the use of correlation.
11	Comparing many means	Murder and Relationship Breakups	Case-control study demonstrating the application of a Chi-square test of association
12	Revision	Methamphetamine and Attention	True randomised experiment demonstrating the use of one-way analysis of variance

Note. [a] Copies of the computer laboratory worksheets can be obtained by emailing the first author. [b] Regular classroom tutorials.

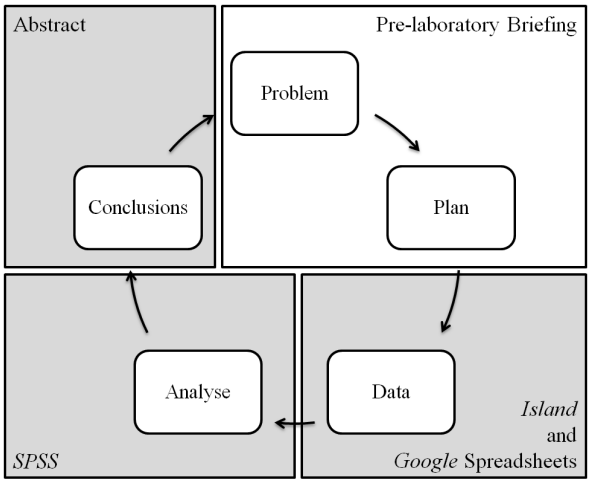


Figure 2. The PPDAC design for each *Island*-based computer laboratory session and the technologies used at each stage. The unshaded block refers to lecture time and shaded blocks refer to computer laboratory time.

Problem and Plan. During the final 15 minutes of each lecture preceding an *Island*-based computer laboratory session, the lead researcher, who developed each session, presented a pre-laboratory brief. The goal of each brief was to introduce the research topic guiding each session, provide an overview of the study design and protocol being used and discuss any important

concepts or considerations. During the briefs, students were instructed to think of their class as a research team commissioned to conduct important psychological research on the *Island* under the watchful supervision their tutors.

Important concepts and considerations discussed during these briefing sessions included sampling, practical data entry strategies and *SPSS* refreshers. Sampling was emphasised during the *Psychological Health Survey* laboratory, which demonstrated recruiting a nationally representative sample from the *Island*. As requiring students to develop and implement a proper random sampling method on the *Island* themselves would take too much time, a compromise was made to have the students randomly select participants from randomly selected suburbs. The process of randomly selecting a suburb from the *Island* was completed by the lead researcher prior to each session. This process was demonstrated in an early brief to provide students with a conceptual understanding of clustered (i.e., suburbs) random sampling. This method of random sampling was used for all other sessions.

Data. Students were provided with worksheets outlining the session's activities, as well as the research protocol discussed in the pre-laboratory briefing. Arriving at the computer laboratory, students would begin by logging into the *Island* and opening the class's *Google* spreadsheet. Class spreadsheets were prepared prior to each laboratory and class access was granted by uploading class email lists. Students were allocated rows identified by their student number in each spreadsheet. The number of rows owned by each student referred to the number of samples they had to recruit for that session. Each row contained the name of a random suburb on the *Island* where a student had to recruit their specified number of Islanders. Students did this using a simple random sampling technique, which involved numbering all eligible Islanders (i.e., only adults) in the suburb and selecting the required number of participants using a random number generator (e.g. http://www.random.org/).

Once the students had selected their random Islanders and obtained consent they would take each one through the research topic's protocol by gathering demographic information and allocating the Islander to complete specific tasks. The student would enter their data recorded from the *Island* in the *Google* spreadsheet. *Google's* cloud collaboration technology enabled students to access, enter and edit the spreadsheet concurrently. According to the *Google* collaboration documentation at the time of the study, there was a limit of 50 simultaneous collaborators. Students were instructed to wait until all others had finished data collection before downloading the class data from *Google* spreadsheets as an *Excel* or *.csv* file. Students were encouraged to help slower students sitting next to them by recruiting further Islanders on their behalf.

Analyse. Once the data were exported from the class's *Google* spreadsheet, the data were imported into *SPSS* and prepared for analysis. This involved data screening and cleaning and defining variable properties. As the data were entered as a class, there were numerous data entry issues; for example, spelling mistakes, data entry errors and ignorance of case sensitivity ensured that students gained valuable experience dealing with messy data. These skills were specifically developed in the *Psychology Health Survey* topic, but were also practiced during other sessions. Students would then use their knowledge gained from the lecture content to perform the appropriate statistical test after checking assumptions. There were no explicit instructions for using *SPSS* as it was demonstrated during the lectures; however, tutors were present to assist students if needed. After conducting the statistical analysis in *SPSS*, students interpreted the results. Students were encouraged to discuss the findings with their peers and tutors.

Conclusions. Following analysis, students were required to write a brief abstract of the laboratory session that included the results and conclusions reached from the statistical analysis. These brief reports were required to be written-up in American Psychology Association (APA) style consistent with the writing conventions of their psychology discipline. Students could request feedback on their abstracts from tutors prior to leaving the session. Completion of abstracts contributed to the students' participation mark.

STUDENT FEEDBACK

In the final week of the semester, 33 students (87%) filled out a short, ethically approved questionnaire, rating their experience of the *Island*-based computer laboratory sessions as well as providing qualitative comments on the positive aspects and areas in need of improvement. In terms of the respondents' demographics, all respondents were domestic students, 2 were part-time, 10 were male and 22 were enrolled in an advanced statistics elective that same semester. It could also be assumed that almost all respondents had completed two compulsory introductory statistics courses during their first year. The average age of respondents was 24 years (SD = 9.69). Students self-reported attending an average of 9/10 tutorial sessions (SD = 1.18) throughout the semester. Students were asked to rate the degree to which their overall experience of the computer laboratory sessions was positive on a 7-point scale ranging from (1) "*strongly disagree*" to (7) "*strongly agree*". The average rating was 6.03 (SD = 0.91) corresponding to an overall positive agreement rate (counting only scores of 5, 6, or 7 as agree) of 94%.

Qualitative comments highlighted the aspects that students valued about the laboratories. Respondents stated the *Island*-based laboratory sessions allowed them to apply and reinforce the concepts covered in the lectures, or as one student stated, "*It was good to not just listen in lectures about what should be done, but to actually do it*". They appreciated the hands on, interactive and practical nature of the tasks, which helped them to develop their understanding of the research process and gain practical *SPSS* skills at the same time: "*Being able to actively practice material that was covered in the lecture, this lead to greater understanding not only of the course material but also how to use SPSS effectively*".

There were a few comments regarding areas needing improvement. Some students stated the sessions were too short given the amount of work required. One student with self-professed, weak computer skills, found the technology and time constraints overwhelming. A few students commented that the time constraints caused them stress and the noise levels in the computer rooms made it difficult to concentrate. A few students were disappointed that they had to wait for the entire class to finish data collection before they could download, import and analyse the data, "*The fact we had to wait for every person to finish data input wasted my time when I could have been doing other things*."

DISCUSSION

Overall, the implementation of the *Island*-based computer laboratory sessions for a research methods course in psychology was highly positively regarded by students who participated in the research methods course. The engaging, relevant and practical design elements of these sessions successfully simulated the data investigative cycle in under two hours with the aid of innovative technologies. However, given the strong statistical background of these students, most with two semesters of statistics knowledge, it is possible that these perceptions were inflated due to the familiarity of the students with the statistical aspects of the course. These sessions need to be evaluated in other student populations that do not have statistical experience.

A few design issues need addressing in future implementations of these laboratory sessions. Some students commented that too much work was required within the two hours. Judging the required amount of time needed for completing the tasks will improve with experience. Being mindful of sample size, a solution to this issue might be reducing the number of samples each student must recruit. Other students with poor or limited technological skills may struggle to stay abreast of the other students in the laboratory. Extra time or assistance might be required for these students, whereas, technologically proficient or stronger students may appreciate the ability to download semi-complete datasets to avoid waiting for slower students; however, these students should be made aware of the implications, e.g. lower statistical power, wider confidence intervals, sample bias, etc.

A number of other general limitations to the sessions must be noted. While the topics covered were interesting to psychology students, the *Island*'s focus on applied human research topics cannot be adapted to all student disciplines that complete research methods courses. Some very important aspects of the data investigative cycle were guided and simplified due to time constraints. For example, stating a research problem and planning a study design was guided by the researcher and the process of random sampling was simplified. The course content and

Island-based laboratory sessions covered only bivariate research design and data analysis. As most real-world research is multivariate, the laboratory sessions described in this report should only be used to build students' fundamental skills, self-efficacy and interest in doing real-world research.

The preliminary outcome of this work is quite promising, especially given the common student perception that quantitative research methods courses are irrelevant and uninteresting. Future research is needed to further evaluate these sessions in other student populations as well as their effect on student learning outcomes.

REFERENCES

Ball, C. T., & Pelco, L. E. (2006). Teaching research methods to undergraduate psychology students using an active cooperative learning approach. *International Journal of Teaching and Learning in Higher Education*, *17*, 147–154.

Barraket, J. (2005). Teaching research method using a student-centred approach? Critical reflections on practice. *Journal of University Teaching & Learning Practice*, *2*. Retrieved from http://ro.uow.edu.au/jutlp/vol2/iss2/3

Bell, B. S., & Kozlowski, S. W. J. (2008). Active learning: Effects of core training design elements on self-regulatory processes, learning, and adaptability. *Journal of Applied Psychology*, *93*, 296–316.

Benson, A., & Blackman, D. (2003). Can research methods ever be interesting? *Active Learning in Higher Education*, *4*, 39–55.

Bridges, G. S., Pershing, J. L., Gillmore, G. M., & Bates, K. A. (1998). Teaching quantitative research methods: A quasi-experimental analysis. *Teaching Sociology*, *26*, 14–28.

Briggs, L. T., Brown, S. E., Gardner, R. B., & Davidson, R. L. (2009). D.RA.MA: An extended conceptualization of student anxiety in criminal justice research methods courses. *Journal of Criminal Justice Education*, *20*, 217–226.

Bulmer, M., & Haladyn, J. K. (2011). Life on an Island: A simulated population to support student projects in statistics. *Technology Innovations in Statistics Education*, *5*. Retrieved from http://escholarship.org/uc/item/2q0740hv

Crull, S. R., & Collins, S. M. (2004). Adapting traditions: Teaching research methods in a large class setting. *Teaching Sociology*, *32*, 206–212.

Gladys, S., Nicholas, Z., & Crispen, B. (2012). Undergraduate students' views on their learning of research methods and statistics (RMS) course: Challenges and alternative strategies. *International Journal of Social Science Tomorrow*, *1*, 1–9.

Meldrum, R. C., & Stults, B. (2012). Student performance in research methods classes: Assessing the predictive validity of the DRAMA scale. *Journal of Criminal Justice Education*, *23*, 381–395.

Saville, B. K. (2008). A guide to teaching research methods in psychology. Malden, MA: Blackwell Publishing.

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*, 223–265