IN-CLASS E-TESTING OF STATISTICS IN A LARGE COHORT OF DIVERSE ABILITY

I.S. Weir*, D.Rh. Gwynllyw and K.L. Henderson

Department of Engineering Design and Mathematics
University of the West of England, Bristol, UK

* iain.weir@uwe.ac.uk

Abstract

We report on our experience of running in-class e-Tests for a new Level 1 Business Decision Making module taken by 800 Business School students at the University of the West of England (UWE Bristol). The module comprises a one semester course covering statistical methodology in far greater depth than had been taught previously and modernised to include analyses using the statistical software SPSS rather than by hand calculations. The syllabus constituted a challenging amount of material to cover, especially since the student cohort was a large, diverse group of non-mathematicians.

The controlled conditions element of the module's assessment comprises the average mark of the best two from three in-class e-Tests. In each e-Test a student receives their own unique data set to work on and performs a complete statistical analysis of it, using SPSS. We describe how we securely assessed the large number of students over fifteen 50 minute PC classes.

The main learning tools were key skills e-Assessments that provided the student with a random data set to import into SPSS to then appropriately analyse and report on. Repeated use of these would ensure that a student would thoroughly learn a key skill and cover various analysis outcome scenarios, for instance significant or not significant test outcomes. Each key skill e-Assessment has multiple embedded links to comprehensive Help pages that provide SPSS 'how-to' information or output interpretation. This form of support is more targeted and immediate than a student having to refer back to the notes. It also encourages the student to self-learn rather than automatically seek staff help. This allowed the staff to concentrate on giving statistical understanding and interpretation advice as opposed to say wasting time on the mechanics of producing SPSS output.

The e-Assessment system used was Dewis which is a fully algorithmic open-source e-Assessment system which was designed and developed at UWE. It is a completely stand-alone web based system used for both summative and formative assessments. Dewis' ability to communicate with the R programming language greatly facilitates the task of generating bespoke student data and providing answers that match SPSS screen output.

Keywords: e-Test, e-Assessment, Statistics, Dewis, R, SPSS

Introduction

We report on our experience of running in-class e-Tests for a new Level 1 Business Decision Making module taken by 800 Business School students at the University of the West of England (UWE Bristol) and partner institutions (Villa College, Maldives, British College Kathmandu, Nepal and Northshore College, Sri Lanka). The module comprises a one semester course covering statistical methodology in far greater depth than had been taught previously and modernised to include analyses using the statistical software SPSS rather than by hand calculations.

The module covers the following statistics topics, which were taught in the order listed:

- Exploratory data analysis including normality testing;
- One sample t-test & nonparametric equivalents;
- Probability & Decision trees;
- Two sample t-test & nonparametric equivalents;
- Critical Path Analysis & Gantt charts;
- Correlation and simple linear regression;
- Chi-squared tests;
- Time series.

The controlled conditions element of the module's assessment comprises the average mark of the best two from three in-class e-Tests. In each e-Test a student receives their own unique data set to work on and performs a complete statistical analysis of it, using SPSS. The e-Assessment system used was Dewis which is a fully algorithmic open-source e-Assessment system which was designed and developed at UWE. It is a completely stand-alone web based system used for both summative and formative assessments (Gwynllyw and Henderson, 2009; Dewis Development Team, 2012). Dewis' ability to communicate with the R programming language greatly facilitates the task of generating bespoke student data and providing answers that match SPSS screen output (Gwynllyw, Weir, and

Henderson, 2015; Weir, Gwynllyw and Henderson, 2015).

The in-class e-Tests are equally spaced throughout the teaching delivery period and prepare the students for the subsequent uncontrolled element of assessment which is a 1200 word report that comprises the statistical analysis using SPSS of a supplied dataset in response to a business question.

Description of the problem or issue

The syllabus presented above constituted a challenging amount of material to cover, especially since the student cohort was a large, diverse group of non-mathematicians. In addition, students were expected to learn how to use SPSS and to gain sufficient competency/understanding of the techniques so that they would be able to perform data analysis when collecting their own primary data in their level 2 research project the following year. Some of the challenges to overcome in designing the way the statistics material was to be taught and assessed were identified as follows:

1. How to teach the students so that they were able to self-learn SPSS;
2. How to securely assess the large number of students over fifteen 50 minute PC classes.

Teaching strategy

The Statistics staff deliver to the students a one hour lecture and a one hour computer practical each week. Business School staff supply relevant context to each of the statistical topics taught through a one hour tutorial per week. The Business School staff were also responsible for the uncontrolled element of the assessment.

The Statistics staff provided notes that covered statistics and use of SPSS and this material was delivered in the lecture. The initial computer practical session concerned the basics of SPSS (e.g. data entry, variable labels, file opening and saving) and the creation of summary statistics, graphics and tables. For this purpose, several bespoke self-learn videos were created so that the students could pause at will and thus work at their own pace. This was deemed necessary due to the wide range of computer abilities amongst the cohort.

From the second computer session onwards, the main learning tools were Dewis key skills e-Assessments. Each use of these provided the student with a random data set to import into SPSS, appropriately analyse and report on. Repeated use of these would ensure that a student would thoroughly learn a key skill and cover various analysis outcome scenarios, for instance significant or not significant test outcomes. Each key skill e-Assessment has multiple embedded links to comprehensive Help pages that provide SPSS 'how-to' information or output interpretation. This form of support is more targeted and immediate than a student having to refer back to the notes. It also encourages the student to self-learn rather than automatically seek staff help. This allowed the staff to concentrate on giving statistical understanding and interpretation advice as opposed to say wasting time on the mechanics of producing SPSS output.

On submission, students are supplied with the correct answer as well as colour-coded feedback on each of their answers. Green indicates that they are correct and red indicates that the answer was incorrect. Access to full and bespoke feedback is important because it has been found that students learn from e-Assessment feedback, using it to perfect their technical knowledge (Greenhow and Gill, 2008).

To supplement the key skills e-Assessments, students are also given pre-prepared word templates which are designed to represent a complete statistical analysis on a specific supplied data set. The templates have the SPSS output removed, numerical values blanked out and inserted multiple choice interpretation decisions to make. Setting it up in this way for students to complete enables them to concentrate on the mechanics of the creation of SPSS output and interpretation of results.

Dewis key skill e-Assessment example

The data scenario for this particular key skill e-Assessment concerns using a 95% confidence interval (CI) for the mean to infer whether the mean IQ of people on a particular medication appears to differ to the general population IQ value of 100. Fig. 1 shows a screenshot of the question that is presented to the student.

The data is supplied to the student via a CSV file as a downloadable link. The successful transfer of the data into SPSS is checked by the student comparing the stated standard error of the mean to that which they obtain in SPSS. Having confirmed agreement, the student is told to input all numerical answers to the same number of decimal places that SPSS reports to.

After the appropriate analysis in SPSS, the student is then required to input the numerical values for the sample size, sample mean and the 95% confidence interval limits before choosing from a dropdown menu the correct inference concerning the problem posed. In Fig. 1 for illustration purposes, the student's attempt has a mixture of correct, incorrect and blank answers.

Fig. 2 demonstrates the initial report which summarises the students attempt and provides an e-Assessment percentage score. Fig. 3 displays 'The Solution' and 'The Report' sections from the feedback of the student's attempt at this e-Assessment. The 'Solution' section supplies the correct answers that the student needed to complete. A pictorial representation of the confidence interval is included along with accompanying text in blue that is extra feedback information that aims to aid the students understanding of the correct inference to the question posed. The 'Report' section indicates, with colour coded marking, what the student has answered correctly (green) or incorrectly (red). It can be seen that the student has one wrong answer and a not answered (NotAns) input.

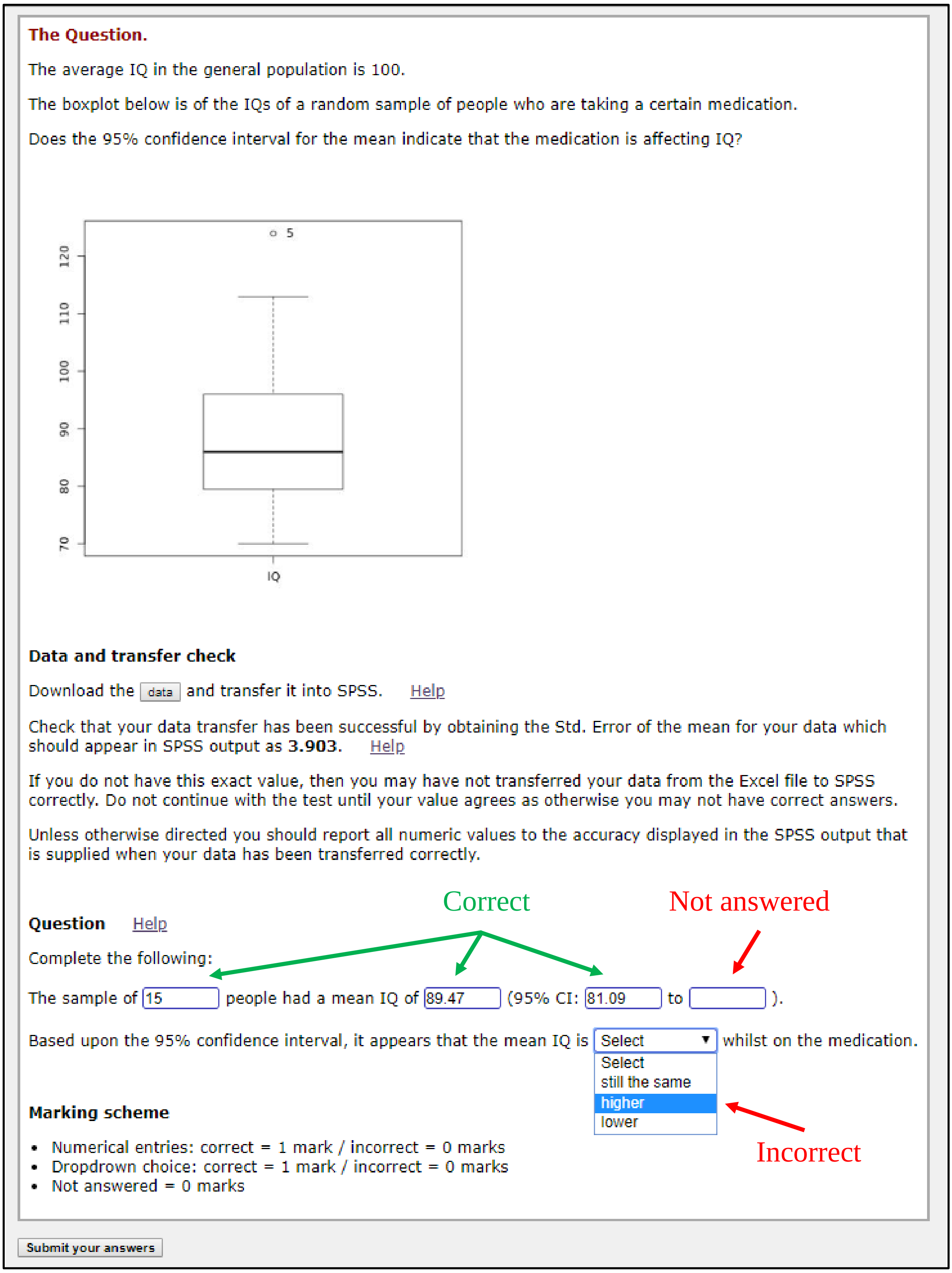

Figure 1. Screenshot of key skill e-Assessment example. The student has entered correct answers bar not supplying the 95% CI upper limit and has an incorrect inference concerning the mean IQ.

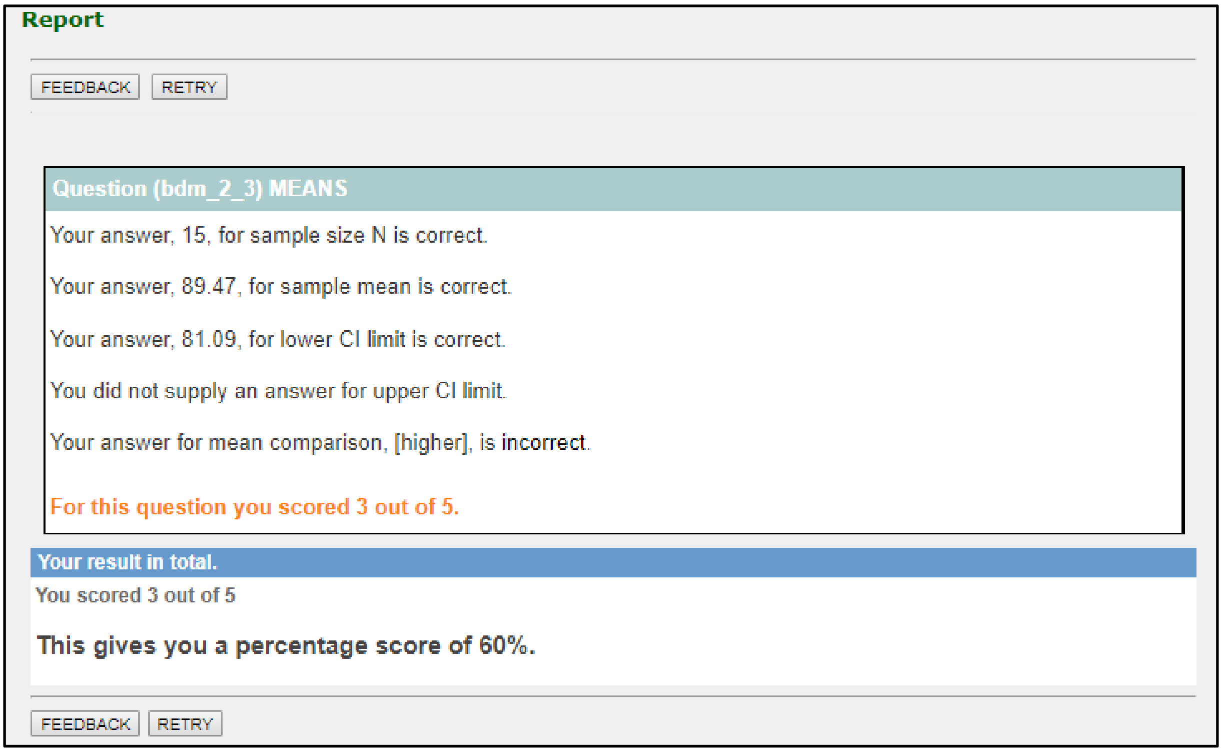

Figure 2. Screenshot of key skill e-Assessment example initial report.

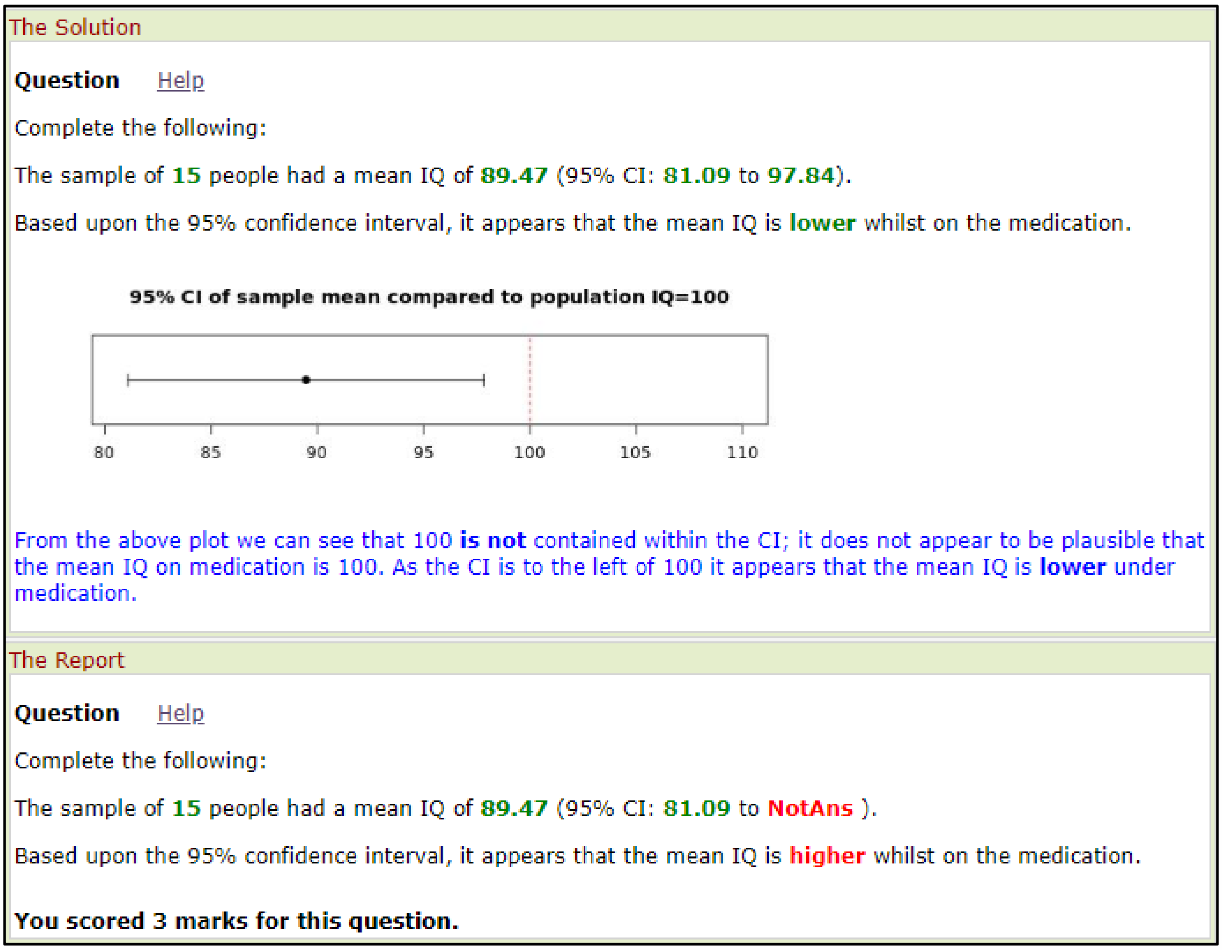

Figure 3. Screenshot of key skill e-Assessment example solution and report sections of the feedback.

The in-class e-Tests

The first in-class e-Test concerns the data dependent choice of the application of either the one-sample t-test or Wilcoxon signed rank test to evaluate a supplier's claimed average sodium content of mineral bottle water. Each run of the test would result in data of random sample size and with a supplier's claimed average that was also random. Furthermore, the data was randomly generated so that with equal chance students would experience both tests and significant or nonsignificant results. The basic statistical analysis tasks tested were the ability to transfer data from CSV format to SPSS, perform an exploratory data analysis for summary statistics, graphics and assumption testing; identification of appropriate statistical test (parametric or nonparametric equivalent), interpretation and reporting of test output. In total there are 22 numerical/dropdown entries that are each assigned 1 mark if correctly answered. There are a further 3 marks available for the correct choice of test to report; this weighting acknowledges the complexity of a decision based upon taught guidelines that consider sample size, skewness and normality.

The second in-class e-Test concerns the data dependent choice of the independent samples t-test, Welch's t-test or Mann-Whitney U test to evaluate the average amount of time visitors spent on two website designs. Each run of the test would result in data of random sample sizes with equal chance that students would experience any of the three tests with significant or nonsignificant results. The statistical tasks required built upon those of the first test and included as extras the interpretation of overlapping confidence intervals and the evaluation of equality of variances. The extra complexity of the guidelines for choosing between three tests was weighted to be worth 4 marks, there were 21 numerical/dropdown entries that are each assigned 1 mark if correctly answered.

The third in-class e-Test concerns the prediction of the sales value of a particular product from the market potential. Each run of the test generates data from a random number of sales territories with varying strengths of positive correlation between the two variables. Students are required to perform and report both a correlation analysis and a regression analysis. It is required that they report on the strength of the correlation, assess the fit of the regression model, report on outliers, interpret the gradient coefficient and produce predictions together with associated 95% confidence intervals. Unlike the analysis in the other two e-Tests, this analysis does not require the student to make any data dependent decisions on techniques to apply. In total there are 25 numerical/dropdown entries that are each assigned 1 mark if correctly answered.

In-class e-Testing procedure

The in-class tests that were run at UWE Bristol were scheduled to run in the same session/room as students' timetabled PC session. This was the only practical way of running the e-tests given the large number of students involved and the fact that the biggest PC lab on campus holds at most 50 students. Each session was scheduled for 50 minutes and the time limit for the in-class e-Test was 30 minutes. Each student was allocated to a group on the University's Virtual Learning Environment (VLE) which corresponded to their PC session. Using adaptive release, access to a practice test for each group was enabled immediately after the corresponding PC class a week before the in-class test. The practice test was exactly the same as the in-class e-Test and remained available to that group until midnight the day before their in-class test was scheduled. During its availability period, students were allowed unlimited attempts at the practice test getting different data (and hence analysis outcomes) each time they attempted it together with full feedback. The Dewis system enforced a strict time limit of 30 minutes to get students used to working under exam conditions.

Using adaptive release the link for the in-class test was only available for a particular PC group during their scheduled computer lab. At the start of the in-class test students were instructed to log into the VLE to access this link and to open the SPSS software. The in-class test link was protected by an examination key and this was different for each PC group. Once all the students were ready to start this examination key was issued and students were directed to enter it and "Start the online exam". This ensured that only students within the room were able to access the in-class test. During the in-class e-Test students were only allowed to access Dewis and the SPSS statistical package. Dewis displayed a persistent grey bottom horizontal bar containing details of the student's identity. This display was intended to facilitate the invigilation process and has been successfully used previously for e-exams run using the Dewis system (Henderson, Gwynllyw and Hooper, 2016). On submission of the assessment, the colour of the horizontal bar changed to pink. Students were allowed to submit their answers ahead of the 30 minute limit and, in such cases, they were instructed to ask an invigilator to view the pink bar before leaving the room. This bar provided the invigilators with an easy visual check of the student engaging with the Dewis system and also of the status of their assessment attempt. The first two in-class tests were invigilated by academic staff involved with the module; however the final in-class test was supervised by staff from the university's exam invigilator team.

Each student's attempt was marked instantly on submission. However disclosure of the mark to the students was delayed until the end of the week. This allowed us time to review and analyse the complete spread of marks once all the students had taken the in-class test. Once we had completed the review process, students were able to log back into the in-class test link and view their submission to get full bespoke feedback for their attempt as well as their mark.

The number of students taking the module at partner institutions was considerably smaller than for those taking it at UWE Bristol. This meant that they were able to give access to the whole cohort in one sitting.

Results

Results have been excellent (e-Test 1: one-sample t-test N = 643, M = 75.2, SD = 20.0; e-Test 2: two-sample t-test N = 640, M = 75.8, SD = 24.0; e-Test 3: correlation and regression N = 380, M = 67.5, SD = 24.7). The decline in the number of students attempting the third e-test is due to the final mark being the best from two; many students with high marks from the first two tests elected not to take the final test. Based upon the best two from the three e-Tests, 83% of students passed this element of the module. This is an excellent pass rate, especially as 9% of students did not sit any of the e-Tests.

Of those students that sat each e-Test, the majority had previously tried the practice tests (e-Test 1: 59%; e-Test 2: 64%; e-Test 3: 54%). Many students made multiple attempts at the practice attempts (e-Test 1: Mdn= 5, IQR = 2 – 9; e-Test 2: Mdn = 6, IQR = 3 – 11; e-Test 3: Mdn = 4, IQR = 2 – 7). Those that did practice experienced a significant uplift in average marks compared to those that did not practice (e-Test 1: 20.8 (95% CI 18.1 – 23.5); e-Test 2: 24.3 (95% CI 20.9 – 27.7); e-Test 3: 20.7 (95% CI 16.2 – 25.3).

There was a high number of students who have viewed their feedback (e-Test 1: 72%; e-Test 2: 70%; e-Test 3: 56%). This is much higher than the 10% that is typical for paper-based work. Research (Race, 2014) shows that feedback has to be quick to be effective, while students still remember clearly the work they were engaged in and online exams are one way of achieving this.

Discussion

Feedback from students to the module has been mixed, in part due to the fact that some students did not feel that there was enough linking between the Statistics and Business elements of the module. Given that 2017/18 was the first year that the module ran in this form this is not totally surprising and greater liaising is planned for future years. However in general the students liked the in-class tests, recognising that "it is fair for all students" because it is not possible to "cheat" as is the case for uncontrolled coursework.

Using e-Assessment for coursework has become standard practice in many institutions (Sangwin, 2013) and it seems likely that online examinations will become standard practice in the near future (Kuikka et al, 2014). However progress has been slower than predicted by Collins, Ripley and Roads (2003). Gray, Sheppard and Ferrell (2016) found that there is considerable interest in online exams yet few reliable sources of good practice. Roads (2016) reports that most organisations expect to increase their use of e-Assessment over the next five years and identified that lack of space for invigilated testing as one potential barrier. In this paper we provide a model for how robust online testing of statistics can be achieved.

The Statistics element of this module is designed to provide a solid foundation for Business School students in using SPSS to perform standard statistical tests. They will experience further material in their second year and will then be required to use the techniques that they have learnt to analyse real data in a substantial research project. It is hoped that the testing regime and the fact that students will have access to the comprehensive materials produced throughout their studies will facilitate their transition to becoming independent researchers.

References

Collins, C., Ripley, M. and Roads, M. (2003) The opportunities for ICT in assessment in Whither Assessment? Edited by C. Richardson, Qualifications and Curriculum Authority.

Dewis Development Team (2012) Dewis welcome page: http://dewis.uwe.ac.uk (accessed 14/06/2018).

Gray, L., Sheppard, M. and Ferrell, G. (2016) Jisc Project: Electronic management of assessment. https://www.jisc.ac.uk/rd/projects/electronic-management-of-assessment (accessed 19/06/2018).

Greenhow, M. and Gill, M. (2008). How effective is feedback in computer-aided assessments? *Learning, Media and Technology* 33(3), pp. 207-220.

Gwynllyw, R. and Henderson, K. (2009). DEWIS: a computer aided assessment system for mathematics and statistics. *CETL-MSOR 2008 Conference Proceedings*, pp. 38-44.

Gwynllyw, R., Weir, I. and Henderson, K. (2015). Using DEWIS and R for multi-staged statistics e-Assessments. *Teaching Mathematics and its Applications*, 35(1), pp. 14-26.

Henderson, K., Gwynllyw, R. and Hooper, A. (2016) Using electronic exams to provide engineering mathematics students with rapid feedback. In: Alpers, B., Dinger, U., Gustafsson, T. and Velichova, D., eds. (2016) 18th SEFI Mathematics Working Group Seminar on Mathematics in Engineering Education Proceedings. SEFI Eur. Soc. for Engineering Education, pp. 105-111.

Kuikkaa, M., Kitolab, M. and Laaksob, M-J. (2014) Challenges when introducing electronic exam. Research in Learning Technology, 22: 22817.

Race, P. (2014) Making Learning Happen, Sage Publications.

Roads, M. (2016) JISC e-Assessment Survey. Final report:http://repository.jisc.ac.uk/6278/14/e-Assessment_survey_report_May_2016_FINAL.pdf (accessed 19/06/2018).

Sangwin, C. (2013) Computer Aided Assessment of Mathematics, Oxford University Press.

Weir, I., Gwynllyw, R. and Henderson, K. (2015). Using technology to inspire and enhance the learning of statistics in a large cohort of diverse ability. In*: IATED, ed. (2015) Edulearn15 Proceedings.*