

# Learning Through Assessment

[Joe Angseesing](#)

*School of Environment, Cheltenham and Gloucester College of Higher Education, Swindon Road, Cheltenham GL50 4AZ, United Kingdom*

## Introduction

Learning and assessment are inextricably linked and this author has, for a number of years, routinely used computer aided assessment (Angseesing 1989, 1998) with assessment as the main objective. This was initially done to facilitate examination marking with moderately large classes, but in the last few years has been used for some coursework as well as for examinations. There are several good reasons for using computer aided marking for coursework, where appropriate:

- it ensures consistency of grading;
- it allows rapid turn round of the assessment so that students have feedback before the next installment of coursework, or before a terminal examination;
- the assessment procedure can be used to force the students to formulate precise answers and to step logically through correct procedures, whereas they may otherwise construct vague conclusions or terminate an exercise with a calculation lacking any context or conclusion; and
- it is easy to obtain a quantitative analysis of where students are going wrong.

Whereas the first two points were the stimulus for introducing computer aided assessment methods in the first place, the last two points have grown to be at least equally important. It is very tedious and perhaps impossible in the time available for a tutor to write a correctly formulated set of hypotheses and conclusions on every script, but easy to program marking software to flag an error of a particular type or to print out a specific comment in response to a specific error. Even more important is the way that computer based questions can teach students to steer a way through technical language, as in the examples introduced below.

## Academic setting

Computer based learning and assessment are used here in a first year module entitled *Environmental Data Handling*. Many of the first year modules in this School are shared by a broad range of clients. Students undertaking *Environmental Data Handling* are primarily Major or Joint Honours students in Physical Geography, Geography, Geology, Natural Resource Management and Environmental Science but there is a scattering of students across other environmental Fields of Study.

This College operates a modular scheme in which undergraduate students undertake eight subject-oriented modules per year for three years; in addition there are two skills based modules for first year students. Most students are on two-subject programs which can be Major-Minor or Joint-Joint. The academic year is divided into two semesters, and normally a student takes four

subject-based modules per semester. Within each semester there are twelve teaching weeks followed by three weeks for revision and examinations. Each module has two or three (the latter for practical subjects such as the sciences) hours contact time per week and can expect another four to six hours of independent work.

In *Environmental Data Handling* contact time consists of a 50-minute lecture followed by a practical computer workshop where students work through examples using printed workbooks with tutors present to resolve problems; the independent work consists partly of reading in preparation for the next session, partly of working through CAL packages and primarily carrying out a weekly task. The weekly problems are assessed, but the coursework is collected in three installments in weeks four, eight and twelve.

The strategy adopted for most standard modules at the science end of the spectrum covered in this School is 50% coursework and 50% examination. Coursework assessments give students practice at constructing coherent reports on the work they have carried out, with a professional standard of layout, including graphs and tables, expected. Examinations ensure that students can work individually; in *Environmental Data Handling* a multi-choice computer based test is used to encourage revision across the entire module syllabus.

## **Curriculum, hardware and software**

*Environmental Data Handling* is an introduction to using computers to produce or carry out: graphs, tables, samples, calculations, descriptive statistics and some standard inferential tests and procedures. This Campus has seventy networked Pentiums and another 25 available as stand-alones for specialist software; the network is College-wide and students can access their own server-based work from other campuses. The machine room is block-booked for the workshop sessions, whereby students learn how to use the necessary software: *Excel*, *SPSS*, *Word*, *Statistics for the Terrified* and *GeographyCAL*. The last two are CAL packages: *Statistics for the Terrified* is an excellent interactive program which explains statistical terminology such as significance, standard error and confidence limits as well as descriptive statistical terms and the procedures for simple tests such as the t-test. *GeographyCAL* runs through the different kinds of sampling regime that can be applied.

Home-produced software is used for both examinations and the part of the coursework which is computer assessed. 50-60 students register for the module each semester, offering great time savings in marking now that the assessment regime has been established.

## **How the coursework assessment works**

Students have to submit a *Word* report on all the practical work carried out during the module. In addition they have to sit at a terminal and respond to questions on many of the tasks carried out - the assessment program is available for several days. The report ensures that they can port graphs and tables from *Excel* and *SPSS* to *Word*, and also ensures that they have carried out their own work; they are warned that no marks are awarded for any computer assessed section if the corresponding section is absent from the hard copy. Some marks are awarded by the tutors - for

some open-ended reports, for graphs and for a clear and properly-labelled layout; 70% of the coursework total is based on the computer mark.

Individual questions are of two basic types: those with an open-ended numerical format and those in a multi-choice format. The open-ended format is used for answers to all calculations and for some other numerical work - such as the values of items in a sample collected from a printed data array using a prescribed method. The multi-choice format is used to make the students identify the best hypothesis, the best conclusion, a reason why a particular statistical test is invalid, the measurement scale of a set of data and so on. An example of each is given below:

What is the value of the Index of Dispersion for the pebble sample in q.9.1?

In q.8.1 which is the best Null Hypothesis for comparing deposit 1 vs. deposit 2?  
Type in only the numerical code of your selection:

1. No difference between the clasts in the two deposits
2. There is a difference between the clasts in the two deposits
3. No difference between the clast angularity in the two deposits
4. There is a difference between the clast angularity in the two deposits
5. No significant difference between the clasts in the two deposits
6. There is a significant difference between the clasts in the two deposits
7. No significant difference between the clast angularity in the two deposits
8. There is a significant difference between the clast angularity in the two deposits

All answers are marked on a scale 0-5. In open format responses 5 marks are given for the correct numerical answer rounded to the correct number of decimal places with some deduction for errors - usually 4 for a rounding error or citing many significant figures, 3 for a very crude rounding error, 2 for the correct order of magnitude only; mark band widths are pre-programmed into the marking package. For some multi-choice format questions the mark allocation is all or none (as in yes/no questions such as: Should you reject the Null Hypothesis?) but some responses allow a vague answer to collect some of the marks. For example, in the array of responses for the Null Hypothesis example above the students are expected to use the phrase significant difference rather than just difference, to identify a Null rather than a Research type of Hypothesis and to pinpoint what feature of the sediments they are comparing. 1, 2 or 3 marks are awarded for responses which are imprecise but which have identified the correct type of hypothesis.

### **Feedback for students and tutors**

Students receive a printout logging their responses to each question and the marks obtained for each. In the past the printout has not recorded precisely why partial marks were given, just what the mark was. Reasons for partial marks were covered in generic comments. An updated version of the program will be used in the forthcoming session, automatically generating a comment for each particular response.

For the tutor there is a breakdown of how many completely correct marks and part-marks were awarded for each question, and of the frequency of distractor selection in the multi-choice questions. The most interesting outcome is that there is evidence for student learning both during the assessment-feedback cycle and during the point at which students select an answer from an array. There is also possible evidence in the opposite direction - for inattention to earlier instructions and results during later tasks.

### **Rounding errors in calculations**

All three coursework assessment stages include calculations which are assessed by input to a computerised assessment program. Some marks are lost due to imprecise rounding even though the correct procedures were used in calculating an answer. The ratio of 5s to partial marks across the three marking points decreased slightly for one student cohort in 2000-2001:

	correct	part-marks	wrong
week 4	454	227	299
week 8	440	240	320
week 12	410	236	304

Twenty questions within each batch of coursework were selected for analysis (there were only 20 in week 4, the first 20 within each of the other batches); when the number of responses in each category are aggregated across the questions the ratio of part-marks to full-marks (part-marks mostly rounding errors) increased slightly. A possible reason is that details on correct rounding and significant figures were provided early in the module, and the weekly instructions were less and less explicit about significant figures in succeeding weeks. It is possible that there is no real trend as this is a (large) sample of the class but not the whole student group in each case, and different students may appear to different degrees in the 'wrong' column, thus different individuals may be represented in the correct versus part-marks breakdown for different questions. No longitudinal data on individual students has been abstracted.

### **Selection of hypotheses - I**

Inferential statistics, including the assessment of significance for correlation coefficients and regression coefficients, are not included until week 7 of the twelve teaching weeks. In week 7 there were two questions, similar to the example above, where students had to select the most appropriate hypothesis from a list of eight alternatives; in weeks 8 to 11 there were four such questions; the analysis of multi-choice format responses is restricted to these questions. An analysis of a batch of students in the same cohort produced the following information (the table shows the aggregated totals for all students and all the 'select a hypothesis' questions):

week	correct	wrong	no response	% correct of
------	---------	-------	-------------	--------------

				responses
7	53	17	12	76
8	83	33	48	72
9	109	19	36	85
10	114	14	36	89
11	132	8	24	94

There is a rising trend for both number of correct responses and proportion of correct responses if null responses are eliminated. The same comments apply as for the previous example on the composition of the groups providing data. Students received feedback in week 10; no formal tuition on phrasing hypotheses was provided in weeks 9-11 other than worked examples given in the practical notes for each week. There are three possible factors that may have contributed to the rising trend, assuming it to be real:

1. increasing familiarity with the technical terminology due to practical notes;
2. the arrival of feedback (but only half of the cohort picked up their assignments and comments before the examination period); and
3. increasing familiarity with the technical terminology due to repeated attempts at questions in a similar format.

The analysis of the summaries of the cohorts computerised assessment does not allow this to be further resolved, except in the case of the final batch of coursework, which was also assessed by a tutor. See discussion below.

### **Selection of hypotheses - II**

For the final batch of coursework the computer-awarded marks for week 11 (regression analysis) were compared with the tutor assessment of the hard copy submitted by the same batch of students. This procedure was very revealing (aggregate of four hypothesis formulations for 35 students, null responses being ignored):

	correct	poorly phrased or incorrect
computer-marked	132	8
tutor-marked	84	56

This time the same group of students was assessed in two ways - according to their response to the multiple choice question in the computer based assessment and according to tutor-judgement whether their hard copy hypothesis statements met all the criteria for full marks - whether they:

(i) were the correct way round (that is the Null Hypothesis was indeed the Null); (ii) included the term 'significantly different' rather than just different; and (iii) specified precisely which coefficient was being tested. The fact that many reports did not meet the full criteria, often having vaguely expressed Null and Information Hypotheses, offers the possibility that confronting the computer-marking program (after the written document had been prepared) was a useful learning experience enabling the students to sharpen imprecise formulations of their own. The possibility arises that there was some cross-stimulation amongst students, but this would not explain the rise in correct computer-responses from weeks 8-11, which were all assessed together in week 12. Further, there was no lecture in week 12 to allow students extra time to complete the last batch of coursework within the timetabled session; more than 70% of the cohort (those included in the analysis here) submitted their answer files within this session with tutors present.

## Conclusions

Several years experience of computer-marking of coursework have convinced the author that working through the assessment program is itself a valuable learning experience, enabling students to focus their ideas by presenting them with a choice of imperfect answers as well as the ideal one. Analysis of the performance of one cohort of students gathered some evidence to support this view, although good practice at carrying out one particular operation does not always 'stick' from one confrontation with a problem to the next.

## References

1. Angseesing, J. P. A. (1989) Open-ended computer-marked tests, *Teaching Earth Sciences*, **14**, 17-19.
2. Angseesing, J. P. A. (1998) Computer-marked tests in Geography and Geology, in D. Charman and A. Elmes (eds), *Computer-based assessment (volume 2): Case Studies in Science and Computing*. SEED Publications, University of Plymouth.

Joe Angseesing  
School of Environment  
Cheltenham and Gloucester College of Higher Education  
Swindon Road  
Cheltenham GL50 4AZ  
United Kingdom  
[jangseesing@chelt.ac.uk](mailto:jangseesing@chelt.ac.uk)