

---

# Pre- and post-testing in evaluating a CAL program: preliminary findings

Jennifer M. Wilson,\* Richard A. Lerski\* and Cathryn Hau\*\*

\*Department of Medical Physics, University of Dundee

\*\*Department of Epidemiology and Public Health, University of Dundee

---

*Pre- and post-testing have been used as summative evaluation tools for the CALRAD (Computer-Assisted Learning for RADIation protection) program. The method by which such testing was implemented is discussed, and the data collected is analysed. One question which is addressed is whether the same or different tests should be used as a pre- and post-test. Conclusions about the use of this evaluation method are presented, along with further considerations which should be taken into account.*

## Introduction

The aim of this paper is to examine how pre- and post-testing were used in the summative evaluation of the CALRAD (Computer-Assisted Learning for RADIation protection) program. Pre- and post-testing refers to the process whereby an assessment test is given before students use educational materials (pre-test) and after the students have used the materials (post-test). Any change in the students' performance can then be attributed to the use of the educational materials, in this case the CALRAD program. The results may provide information as to whether the program is educationally effective.

## The CALRAD program

The CALRAD program was developed with Teaching and Learning Technology Programme (TLTP) funding. It aims to teach the principles of radiation protection to undergraduate medical students and postgraduate medical staff. The program has been approved for such teaching by the Institute of Physics and Engineering in Medicine.

The emphasis was on producing a teaching package which would be an improvement on the current method of teaching the course (a six-hour lecture course on radiation protection), and which would be more satisfactory to both students and teaching staff.

CALRAD was designed to allow students to direct their own learning, by providing them with aims and objectives, self-assessment questions and summaries in each of the eight

---

sections that are covered in the program. These sections include Basic Physics of Radiation, Biological Effects, and The Use of X-Rays. Hypertext facilities were used so that the students could study particular topics to the depth they required. The program incorporates multimedia in the form of video clips and animations to illustrate dynamic processes such as radionuclide kidney scans. Problem-based learning, in the form of four case studies, has been used to give the students an opportunity to apply their knowledge in the context of real-life situations (Wilson, 1995). The case studies give the students an opportunity to learn by doing or to learn from experience, as well as the opportunity to test their knowledge. A paper-based Student Study Guide accompanies the program, containing a summary of the content and a glossary of terms. The students can keep this Study Guide for future reference.

### **Summative evaluation**

In order to discover whether the program enables the students to learn successfully, a summative evaluation was undertaken. A summative evaluation normally looks at the summed effects of a program once it has been fully implemented in its educational context (Universities Funding Council, 1992), that is, when it is being used by the target audience in the environment for which it was intended.

The evaluation methods which were used in the summative evaluation of CALRAD were:

- pre- and post-tests to measure educational effectiveness;
- attitude scales in a questionnaire;
- open questions in a questionnaire;
- computerized logging to determine how students were using the program;
- observations and discussions.

This list includes a variety of complementary evaluation methods – both open-ended methods for detecting unexpected results and fixed methods for generating comparative data that can answer specific questions (Draper, 1994). However, for the purposes of this paper, only the use of pre- and post-testing will be examined.

### **Method of using pre- and post-testing**

The CALRAD program was used to teach the radiation protection course at the University of Dundee in October 1996. A total of 140 third-year medical students were timetabled in four classes of around 35 students to attend for two 2-hour sessions to use CALRAD. The students were divided into the four classes alphabetically.

There are currently no facilities in place to examine the subject, and in order to receive a completion certificate, the students were required to spend between 3.5 and 4 hours on the program. Prior to using the program, the students were given a pre-test, and on completion of the program, a post-test. They were made aware that these tests were to evaluate the program and not to evaluate individual students (Draper *et al*, 1994). The tests were designed to check that the students had achieved certain important learning objectives on completion of the program, and in this way the educational effectiveness of the program could be tested.

Out of the class of 140 students 111 completed both a pre-test and a post-test. There were various reasons why the other 29 students in the class did not complete both tests, including absence, and the fact that some had already started to use CALRAD before the first session.

It was initially felt that pre- and post-tests could be used in order to measure the 'learning gain' of the students. However, it was realized that it would be difficult to create a test which would give a true reflection of how many of the large number of objectives a student actually achieved before and after completing the program. The results obtained very much depend upon which questions are actually asked.

The tests were designed such that they would not take long to complete. Each test was marked out of 15 points. The questions were divided into two sections:

- Four questions which were intended to be of Higher Grade or A-Level Physics standard. Their purpose was basically to give the students some questions which they could answer in the pre-test. The results for these questions will not be presented here.
- Seven questions, with 11 marks available in total, which were intended to be easy to answer after completing CALRAD. The students' general knowledge of medicine and physics may mean that they may get some of these questions correct in the pre-test, but it was felt that they should be scoring highly on these particular questions in the post-test. It is the scores on these questions which are marked out of 11 which have been analysed for this evaluation

The questions were designed in such a way that it would be difficult to guess the correct answer. They were mostly designed as short-answer questions, with very few multiple-choice style questions being used. This was to ensure that any questions which were answered correctly could be attributed to the students definitely knowing the answers.

Two tests, A and B were created, intended to be of equivalent difficulty. The four classes were given one of the two tests (A or B) as a pre-test before using CALRAD, and one of the two tests (A or B) as a post-test after using it, as follows:

Monday class	Pre-test A	Post-test A	(20 students completed both tests)
Tuesday class	Pre-test B	Post-test B	(32 students completed both tests)
Thursday class	Pre-test A	Post-test B	(34 students completed both tests)
Friday class	Pre-test B	Post-test A	(25 students completed both tests)

Some classes were given the same pre- and post-tests, and some were given different pre- and post-tests in order to analyse which of the following two opposing views might be correct:

*It is better to give the same pre- and post-test*

The same pre- and post-test should be used for testing students, as this is the only way that the evaluator can be certain that the tests are equivalent. If the students are tested on the same questions, any improvement in the number of questions which are

answered correctly in the post-test can be definitely attributed to the students having increased their knowledge as a result of using the program; as Draper *et al* (1994) have put it: 'Having identical tests may be boring for students, but it makes us more confident of having an exact comparison between tests.'

*Students should not be given the same pre- and post-test*

It may not be a good idea to use the same pre- and post-test as students may spot the answers to questions which they failed to answer in the pre-test while they are working through the program. These answers may then take on a significance which they would not normally have if they had not completed the pre-test. Students may remember them, and if the same test were given again, the students might then answer these questions correctly.

It was felt that both of these seemed to be sensible points of view. The scores that the students achieved in the pre- and post-tests were then analysed to discover firstly if there had been a learning gain, and secondly to discover if the fact that they were given the same or a different test as a pre- and post-test had any effect.

## Results

Non-parametric statistical tests were used as the sample of students was small, and the results were found to be non-normally distributed (Bland, 1995).

For all 111 students who completed both a pre- and a post-test, scores were obtained for the questions marked out of 11 (i.e. those which intended to be easy to answer *after* completing CALRAD). When the post-test and pre-test results are compared (Figure 1), it can be seen that there appears to be a definite improvement in the scores achieved, with the students' median score being 4 out of 11 in the pre-test and 10 out of 11 in the post-test.

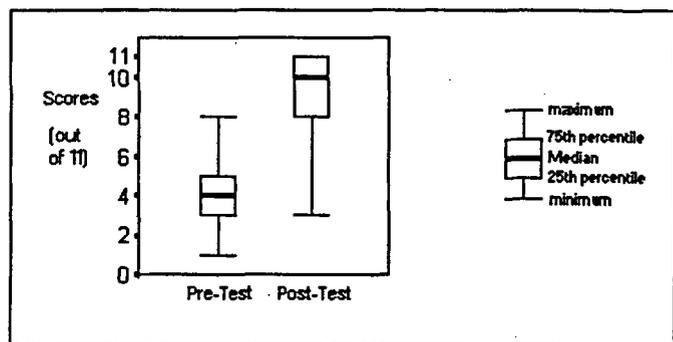


Figure 1: Pre- and post-test results

Of these 111 students, 105 did better in the post-test than in the pre-test, five students achieved the same result, and only one did worse in the post-test. Using the Wilcoxon test (Bland, 1995), the learning gain was found to be highly significant ( $p < 0.0001$ ).

The results for the individual days were then examined to see if there was any effect depending on whether the students were given the same or a different pre- and post-test. The four sets of results (i.e. from the four separate classes) were analysed separately to look at the learning gains. These results are displayed in Figure 2 which illustrates that the effect

of the intervention (teaching with CALRAD) was so great that the effect of being given the same or a different pre- and post-test appears to be relatively unimportant. However, it is not possible to put any statistical interpretation on the effects of this due to the small sample size. It would be necessary to repeat the experiment with more than four classes, as the results may depend on the particular learning environment in each class.

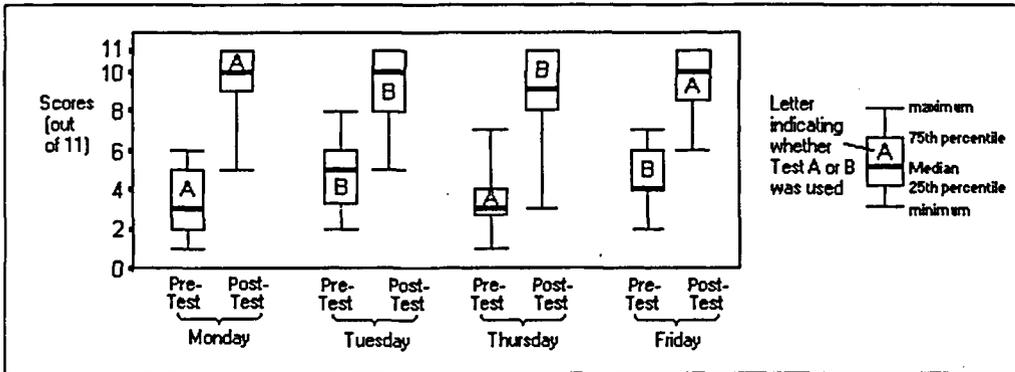


Figure 2 : Graph of results for individual days

The results for the individual days were then examined to discover if there was a significant improvement in the performance of the learners for each day. Using the Wilcoxon test, the learning gain was highly significant for each day ( $p < 0.0001$ ). It was therefore concluded that the CALRAD program had provided an effective educational experience for the students.

### Further considerations

There are a few aspects about the use of pre- and post-testing which the evaluator should be aware of:

- It is fairly difficult to design two tests which are of an equivalent difficulty level. In this instance, the sample is too small to determine whether tests A and B were of similar difficulty.
- It should be remembered that the pre-test and post-tests may in themselves be a factor in promoting learning by reinforcement and practice, and that in fact what is actually being measured is the combined effects of the learning materials plus the tests (Draper *et al*, 1994).
- For a long time, good practice has called for delayed post-tests. It may be that students do not assimilate all they have learned from a course until they have gone away and revised for their examinations. It may therefore be more sensible to carry out a post-test after a suitable delay. This would also mean that a measure could be obtained of how much information the students had retained after a period of time. However, the problem with this is that in the intervening period, the students may have done something else relevant, for example reading a text-book or completing an assignment, and so the measurement of learning gain will then be confounded (Draper *et al*, 1994).

An immediate post-test, such as the one that has been done for this summative evaluation, will pick out the effects of a specific educational intervention, for example CALRAD, even if important aspects of learning and reflection take place only later (Draper *et al*, 1995).

## Conclusions

The use of pre- and post-testing was one of the most valuable evaluation tools used as part of the summative evaluation process for the CALRAD program. The main strength of using pre- and post-testing is that the students' performance before and after an educational event can be measured, and this leads to a determination of the educational effectiveness of that event. In the case of CALRAD, the fact that the students' learning gain was found to be highly significant shows that the program had been educationally effective.

Some investigations were also carried out to discover if the same or different pre- and post-tests should be given. In the case of CALRAD, the effect of the intervention (teaching with CALRAD) was so great that the effect of being given the same or a different pre- and post-test appeared to be relatively unimportant. However, the results should be treated with caution due to the small sample size.

## References

- Bland, M. (1995), *An Introduction to Medical Statistics*, Oxford: OUP.
- Draper, S. W. (1994), 'Evaluating courseware' in *Information Handbook Version 2*, Learning Technology Dissemination Initiative, Heriot-Watt University, 56-60.
- Draper, S. W., Brown, M. I., Edgerton, E., Henderson, F. P., McAteer, E., Smith, E. D. and Watt, H. D. (1994), *Observing and Measuring the Performance of Educational Technology*, a report by the University of Glasgow's institutional project in the TLTP Programme 'Teaching with Independent Learning Technologies', Robert Clark Centre for Technological Education, University of Glasgow.
- Draper, S. W., Brown, M. I., Henderson, F. P. and McAteer, E. (1995), *Integrative Evaluation: An Emerging Role for Classroom Studies of CAL*, TILT report, University of Glasgow (derived from a paper at CAL95).
- Wilson, J. (1995), 'TLTP medical physics project: an outline of progress so far', *CTICM Update*, 6 (1), 3-5.