Learning Through Assessment:
OLAL .. On-Line Assessment and Learning

Phil Davies
Senior Lecturer
School of Computing
University of Glamorgan

Abstract

This paper describes the development of a multiple choice, computerised, and formative **O**n **L**ine **A**ssessment & **L**earning System (**OLAL**), and the results and feedback with respect to the system's introduction into a degree module as a means of assessment, and as a tool to facilitate learning.

This system was developed to replace the current exam based process as a means of assessing a level two degree module in Computer Communications and Networks within the School of Computing at the University of Glamorgan.

The **OLAL** system, significantly reduces the marking load on the academic member of staff, but at the same time clearly addresses the issues of instantaneous student feedback, and the maintaining of standards within the subject area without the final examination process being undertaken. It allows the assessment of the student's understanding of the subject area, rather than merely addressing their ability in report writing or proficiency in the use of a specific development tool.

The use of **OLAL**, permits the cumulative assessment of the students at four stages throughout the module. The students, whilst performing the test, are permitted two attempts, with the same questions being presented on each occasion. The cumulative score of these passes, in the ratio of 60/40, is then stored away to the student database for assessment purposes. The test includes a timer per question, ensuring that the students are in a "pressure environment" whilst performing the test. It also allows the student time to make use of the formative nature of the system in checking whether they were correct, or if not, then why? The marking scheme of the system includes a score of plus two for a correct solution and a punitive measure of minus one for an incorrect solution. This simplistic measure of allowing two attempts of the test provides a very effective learning mechanism, rather than a pure assessment exercise, as is the normal situation in the standard non-formative examination process.

The test also provides a valuable tool in automatically reporting on questions which have caused problems to the students, and hence suggests areas of questioning to be passed through to the next stages of assessment, and also indicates areas of the course which may require further lecturing or directed support.

The paper analyses the overall results produced by the introduction of this system with respect to the students' perceived understanding of the subject area. It also addresses the efficiency of the system, and also reports upon the student feedback concerning their acceptance of the system as both a means of assessment and its formative and learning features.

Introduction

One of the major problems facing lecturers in higher education, is the problem of the ever-increasing numbers of students enrolling, and the formal assessment of these students. Staff / Student ratios are increasing across the higher education sector. It is not uncommon for lectures of over two hundred students per room, with only the **SSL&L** (**s**it down, **s**hut up, **l**isten and **l**earn) method of lecturing being available. One of the major problems with this method is how do lecturers assess throughout the course, whether the students have actually understood the material that is being presented? This quantitative problem needs to be addressed so that it does not have a detrimental affect on the qualitative process of assessment. The use of CAA (Computer Assisted Assessments) is not a new one, however, the acceptance of such systems has often produced a note of contention amongst staff, with respect to the "dumbing down" of the assessment process. This paper addresses the issues with respect to the production of such a CAA system and also presents student feedback concerning the benefits of using such a system. An analysis of the learning benefits of utilising a simple re-presentation of the assessment process is also covered.

Background

Within the higher education sector, there has tended to be a real problem with students of a variety of backgrounds and experience, being able to understand the often difficult architectural concepts of computing. Finding a method that addresses the assessment of such modules has been a constant problem, without introducing the ancillary problems such as prior knowledge of programming, knowledge of tools, etc. The previous assessment process for the level two module in Computer Communications and Networks in the School of Computing at the University of Glamorgan, comprised of a 50/50 split between examination and coursework elements. The end of semester results were often disappointing.

Examination Process

One of the major drawbacks of an examination process, is whether the key points of a module are being fully addressed. A slightly cynical example would be that a module comprises of twenty key points. We produce five examination questions that might suggest that only 20% of the module content is being addressed. The rubric normally would allow a choice of three out of five questions, i.e. 15% of the module content being assessed. To pass this element of the assessment process, the student normally needs to achieve a score of 40% i.e. the student does not know 60% of the 15% of the module content he/she is being assessed on, and still passes! Therefore, how

can we possibly as lecturers, "hand on heart", present such results and confidently pass on these students to further study, where this very module may be a prerequisite of a module further on down the line.

Coursework

The alternative to relying solely on this examination process is to introduce a coursework element to the module. Often this coursework will require a programming or report type element to their work. However, will we be assessing the student's knowledge of programming or their ability in English, rather than assessing their knowledge of the module in question. There is also the problem of the **EHFF** syndrome. This is the **E**xcessive **H**elp **F**rom **F**riends syndrome that is prevalent in the assessment process. We must be able to ensure that the work produced is the sole effort of the student in question, rather than an amalgamation of the efforts of a number of friends. This problem is fairly easy to address with small groups, however, when attempting to monitor this problem with a class of over one hundred students it is daunting task, which again as lecturers we are not able to fully guarantee. The collation of submitted courseworks by known groups of associates has been an aid in the past, however, with total anonymity being introduced, this may well result in uncontrollable mass plagiarism. One of the major benefits of assessing throughout the progress of a module is that students are provided with a formative feedback (however with numbers increasing, the time delay in the return of courseworks is increasing, thus making this advantage minimal).

Therefore, it was decided that a new method of assessment would be introduced as part of the assessment process for the Computer Communications and Networks module. The system introduced was a conventional multiple choice type system, comprising of a question stem, a correct answer and two other distracter answers. This type of system would be ideal for assessing the student's declarative knowledge, but had little effect on assessing their procedural knowledge (Snow & Lohman). This testing would contribute 60% towards the module's subject mark, and would attempt to amalgamate the advantages of the isolation presented within the examination process, plus the continual assessment and formative nature of courseworks. This would hopefully ensure that the students would learn from their mistakes. A written assessment system could have been introduced that would have required at least three or four examination sessions throughout the semester, with immediate feedback. However, this was not feasible as marking over a hundred scripts, and providing immediate feedback seemed to be impracticable (and an extremely daunting task).

The Development of OLAL (On-Line Assessment and Learning)

The development of a computerized computer system seemed to provide a possible solution to the problems of marking. The use of multiple choice questions appeared to provide a "simple" method of assessing the student's knowledge and understanding of the module content. The system decided upon was a selected response type system rather than a constructive one,

mainly due to ease of development. At various stages throughout the course, the students during their practical/tutorial sessions could be presented with this computerized test. The results of this test could then be stored away automatically by the system, and also feedback could be provided to the students as they sat the test. This would mean that the students would be continually assessed (i.e. they would be forced to study the material throughout the course, and not just rely upon cramming the night before the examination), and also receive feedback concerning their understanding of the subject area. It was decided to produce a custom made application in Visual Basic, rather than utilise an off the shelf package, so that the full requirements of the system could be met.

A list of the perceived requirements of the test are presented below:

a) An introductory part to the system that would explain the system's operation and allocation of marks
b) A method of recording results against a particular candidate
c) A method of question presentation and selection
d) A time limitation to simulate an examination process.
e) A suitable marking system, including some weighting mechanisms.
f) The inclusion of some form of formative and learning mechanisms.
g) Some method of feedback to the lecturer indicating areas of problems within the test

a) An introductory part to the system that would explain the system's operation and allocation of marks

The initial screen would display a list of instructions informing the student how the test would operate, how they would gain or lose marks, etc. Once they had started the test, they would have to complete it, otherwise a zero score would be recorded. This would stop the student from starting a test, getting something wrong, and then starting again to get it right the second time.

b) A method of recording results against a particular candidate

When the student pressed the button on the screen indicating that they understood the rules, they would be requested to select their name from the class list. When they had selected their name, the box would expand to make the name easily readable. This would provide a bonus, in that the lecturer would be able to put a name to a face in the class, which is an important aspect of lecturing, however this is not supported by the increasing movement towards anonymity. This selection of the name would allow the system to automatically record the results of the test at the end.

c) A method of question presentation and the selection

When the students pressed the question button on the screen, a question would appear with the three possible alternative answers. The student would then select what they believed to be the correct answer.

d) A time limitation to simulate an examination process.

Within the test, in order to place the students under a situation similar to an examination, each question would be given a time limit of thirty seconds, which could be indicated by a down counter on the screen. This timing constraint would also provide a competitive nature to the testing to stimulate the students.

e) A suitable marking system, including some weighting mechanisms.

One of the problems of multiple choice type tests is the student just randomly guessing the answers. The use of a confidence type of weighting was considered (Gardner-Medwin), for the creation of a summative mark, however it was decided to keep the system as simple as possible. There was however a definite requirement for some form of punitive measure for the student in getting the answer incorrect. A marking scheme was decided upon that would add two marks for a correct selection, and deduct one mark for an incorrect selection, i.e. minus three for getting it wrong.

f) The inclusion of some form of formative and learning mechanisms.

If the students were incorrect in their selection, the machine would beep to inform them of the fact. This would produce a certain amount of peer pressure, in that no member of the class was keen for the other members to note that they were getting it wrong. It was decided that this use of peer pressure could produce an incentive for the students to put in the effort in revising prior to the assessed tests. After each question, the correct answer was shown. The students then would have as long as they wished to consider this answer, before pressing for the next question. The students on completion of the test, would then be allowed to go through it a second time. This would encourage the students to ensure that they spent time in trying to understand the solution to a question that they might have had wrong the first time through. In passing the marks through for assessment purposes, 60% of the mark from the first pass and 40% of the mark from the second pass would be stored away. This would provide the students with an incentive to concentrate more on the test, especially in going through it the first time. This method of allowing the students to acquire marks for answering questions that they have already seen could be frowned upon, however it is not really any different to presenting students with as seen examination papers.

g) Some method of feedback to the lecturer indicating areas of problems within the test

At the completion of the test, the student's mark would be automatically stored away to the central database for future analysis. Also built into the system would be a method whereby the identification of questions that had caused the class problems, so that analysis of possible areas of student lack of understanding could be quickly identified and acted upon.

Creation of Questions:

There are numerous papers, texts and commercial sources available in helping to develop guidelines for the creation of multiple choice questions, namely (Field & Chandler), (Carneson, Delpierre & Masters), (Dempster), (Haladnya & Downing), (Haladyna T.M).

In developing the multiple choice questions certain rules were decided upon and followed:

- Initially clearly define the objectives and scope of each test in order to meet the learning outcomes
- Ensure that the depth and complexity of the questions clearly reflects the knowledge expectations of the students
- Each question must include a stem and three distracters
- The stem must be presented in a clear, simple and concise manner, avoiding any possible misunderstanding
- The distracters must all be plausible
- Only one answer is unambiguously correct
- Each item is independent of all others. Clues must not be provided to subsequent questions
- Avoid using clues to the correct option
    - Length: do not make the correct choice the longest option
    - Do not provide the correct option as the one with the most detail
    - Grammatical: All options making a grammatically correct statement
- Ensure that the time allocated for each question provides sufficient time for the student to read all of the distracters
- Minimize reading time where possible
- Avoid negative formats in both stem and distracters
- Avoid opinion based answers
- Try to avoid question overlap
- The position of the correct answer should vary and should be random
- Attempt to eliminate guess work in selecting the correct answer by the format of the distracter

This was undoubtedly one of the hardest, and most time consuming aspects of developing the **OLAL** system. The creation of the stem part of the question was relatively easy, based upon past tutorial and examination questions. The production of the distracters conforming to the above rules, proved an extremely difficult and time consuming task. Therefore, the claim of reducing the workload by the introduction of a multiple choice question system does not ring true. Each test comprised of at least forty questions, however, once these questions had been created, then they could be used again in the future, either in subsequent years, or in future tests within the year.

Introduction of OLAL

The test developed was fairly simplistic in appearance. However, it provided a method of CAA and also provided a means of formative assessment. The test was presented on four separate occasions at weeks three, six, eight and twelve. The ratio of the marks passed through to the summative assessment was 5%, 15%, 15% and 25% of the total subject mark of the 60% that had been allocated for this method of assessment. The assessment was cumulative in nature, in that each of the tests included questions relating to all of the work covered up to that point in the module's progress. At the end of the module, a questionnaire was issued to the students, requesting feedback on this method of assessment.

Student Feedback:

Below are included a sample of the student questionnaire responses. I have attempted to produce a fair and balanced representation of the comments.

<u>The students were asked how they had found the OLAL system with respect to its use as a means of assessment:</u>
“Excellent way of gaining knowledge”
“Very Good, promotes learning of a subject early on”
“Very Good”
"better than sex"
“because it required the student to study all aspects of the topic in case they came up”
“doing it twice helped learn better”
“more relaxed, but also hard”
“enjoyable, did not feel like an assessment”
“unfair”

<u>The students were asked how they had found the OLAL system as a method that facilitated the learning of the material within the module</u>
“Brilliant”
“difficult to learn all materials, takes a long time”
“Good”
“Not good, Should be more time to think of answers”
“Second test actually allowed us to learn as we were doing tests”
“good assistant to learning”
“difficult”
“encourages revision”
“good, mistakes were invariably remembered”
“unrealistic”

<u>The students were asked were they happy with the OLAL system providing 60% towards the overall subject mark</u>
“Good”
“Keep Same”
“weighted Less”

“Should have an end of year exam worth 75%”
“Should be worth more 75% - 80%”
“worth 40%”

<u>The students were then asked for any general comments on the OLAL system</u>
“good idea, because we are always learning”
“a good learning tool”
“I liked the way that the tests increased in marks”
“a good test of factual knowledge, conducted in a relaxed environment”
“a practice test to get used to them would have helped”
“try to improve the interface”
“pity other lecturers haven’t thought if it”
“I found this module very interesting and fun to learn”
“Good continuous assessment”
“The structure of the tests could be improved”
“too hard”
“the colour scheme is awful, apart from that very good”
“forces you to remember the answers, not understand”
“encouraged learning”

The overall feedback concerning the tests was very positive, both as a method of assessment and also as a method which facilitated learning. The students in general felt that being tested in an informal environment was more relaxing than having to sit through the normal examination process. One comment was of note, “forces you to remember answers, not understand”. This comment indicates the major problem I feel that is associated with this second pass of the test. Have the students really understood the correct solution or merely remembered it? If this were the case, then it would go against the use of this tool as a learning aid. However the average results produced from test to test, would be expected to remain constant, as questions were passed through from test to test. This obviously required further study and analysis of the test results.

Test Results:

The average score for the class in each of the tests is presented below:

	PASS ONE	PASS TWO
TEST ONE (Week 3)	50.86%	71.58%
TEST TWO (Week 6)	42.50%	73.36%
TEST THREE (Week 8)	44.88%	79.07%
TEST FOUR (Week 12)	55.29%	81.22%

Figure 1

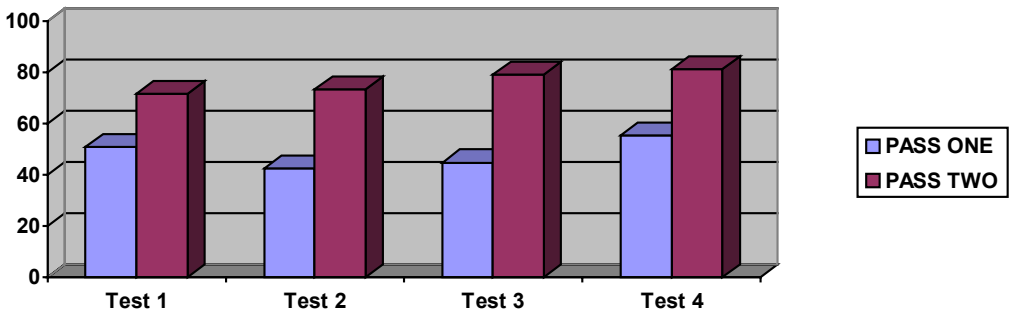

Figure 1 indicates that there was a significant improvement in the percentage scores from the first to the second pass of all of the tests. There was a drop in the average performance for the second test compared with the first test, but this was followed by an improvement in subsequent tests. It should be noted that the second attempt results show a marked improvement from test one through to test four.

As stated previously in this paper, the ratio of 60/40 for first and second pass was decided at the start of the testing. The table below illustrates the affect of changing this split to 50/50, 70/30 or 80/20. As can be seen, there would have been a swing in the final average mark of approximately 10% between a 50/50 split and 80/20 split. On reflection, the 60/40 or 70/30 splits appear to provide the better ratios, taking into account that one of the main reasons of producing the test in its current format is to encourage the students to concentrate and make use of the learning aspects of the testing. This could possibly be lost if they had little to gain from attaining a good result from the second pass (as is the case with an 80/20 split), yet will still reward the students for their initial knowledge and efforts in preparing for the test (which would not be the case with a 50/50 split).

	50/50	60/40	70/30	80/20	RATIO
TEST 1 (Week 3)	61.22	59.12	57.08	55.00	0.05
TEST 2 (Week 6)	57.93	54.84	51.76	48.67	0.15
TEST 3 (Week 8)	61.97	58.56	55.14	51.72	0.15
TEST 4 (Week 12)	68.25	65.66	63.07	60.48	0.25
Average Produced	63.51%	60.73%	57.78%	54.88%	

The ratio of 0.05, 0.15, 0,15, and 0.25 of the tests 1,2,3&4 respectively was decided upon in making the 60% of the subject mark of the module. This was done in an attempt to take into account the quantity and complexity of the work to be covered at different stages of the module. The first three weeks of the module were fairly introductory, moving on to the full module content and complexity up at week twelve. If the breakdown of the ratio had been based directly upon the number of weeks covered, then the average mark, with a 60/40 split, would have been 60.74% compared with the 60.73% as produced. This difference is obviously of no real consequence with respect to the overall results.

Analysis of the Results

The following tables illustrate the frequency distribution of the number of students attaining the average percentage marks, after each of the four tests (non-attendance counts as a zero mark):

Test 1

<10	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90>
1	1	6	6	24	24	27	17	10	0

Test 2

<10	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90>
2	2	3	8	30	30	19	13	8	1

Test 3

<10	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90>
0	0	1	10	19	29	32	17	4	4

Test 4

<10	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90>
7	1	1	5	5	22	32	19	17	7

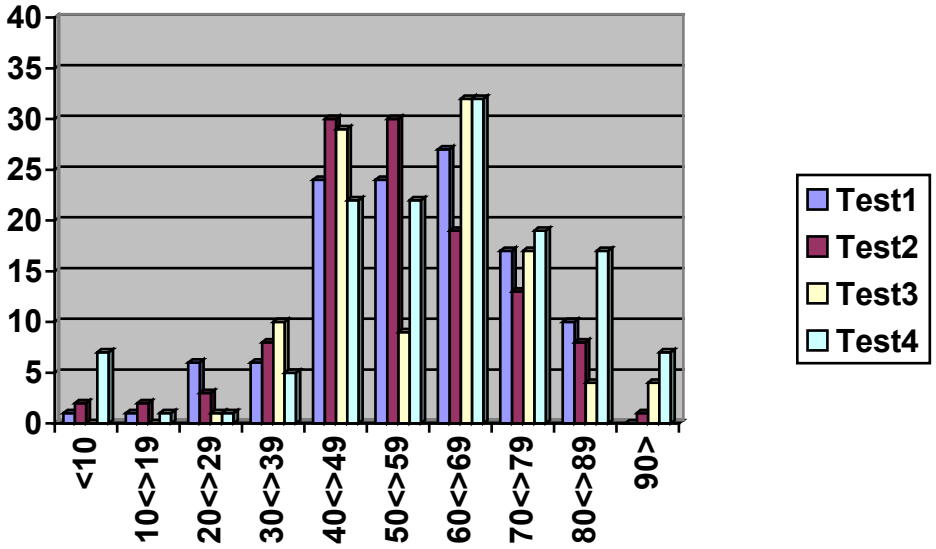

Figure 2

Figure 2 illustrates a good standard deviation of the results, with the majority of the students being found with average scores between 40% and 80%.

The overall frequency distribution of the mean of the results produced after the four tests, based upon a 60/40 ratio of first to second pass of results, is shown below:

<10	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90>
0	1	4	8	8	32	36	18	8	1

Figure 3 shows that based only upon the results produced by the **OLAL** system, out of the total class of 116 students, 103 of the students would have achieved a pass mark of 40% or more.

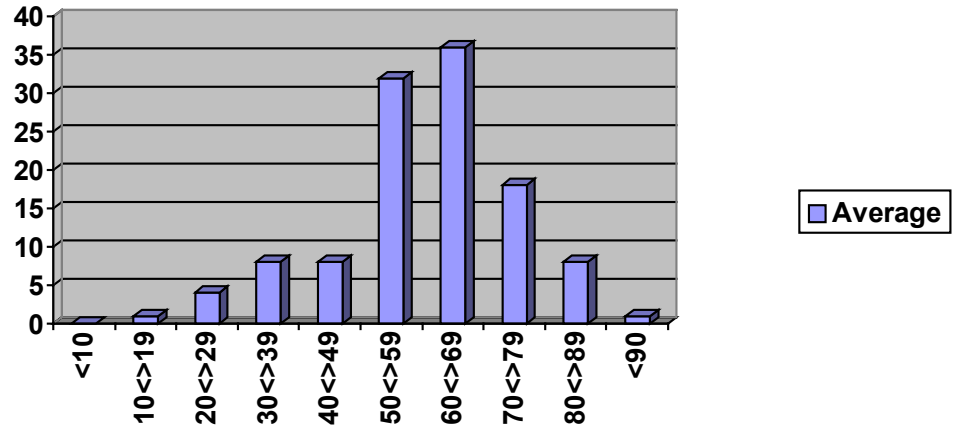

Figure 3

Percentage of Students Passing the Tests:

Percentage of Students Passing		Percentage Passing (Removing Student Non-Attempts)	
Test 1 102/116	87.93%	Test 1 102/115	88.69%
Test 2 101/116	87.06%	Test 2 101/115	87.83%
Test 3 105/116	90.51%	Test 3 105/116	90.52%
Test 4 102/116	87.93%	Test 4 102/109	93.58%

This shows a high percentage of students passing the tests, especially the final test, which in effect should be the hardest test as it covers all of the module's material. The results are an undoubted success. The students have achieved a consistency of performance throughout the tests

Having viewed the basic results, and the changes in the average results throughout the tests, it is noticeable that overall there is an improvement throughout the course of the assessment process. It will be interesting to assess which students have gained the most throughout the module. An analysis was carried out which grouped the students into the categories of percentile average achieved from test one. Having set up the initial grouping

of students, a comparison is carried out for them as they have progressed throughout the subsequent tests.

Initial Groupings from First Test (116 students)

<10	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90>
1	1	6	6	24	24	27	17	10	0

To produce a valid comparison, the students who have missed one or more of the tests throughout the progress of the module have been removed, so that the results indicate the actual marks achieved.

Grouping from First Test (Non-Submissions /Attendance Removed) (108 students)

<10	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90>
0	1	5	5	20	23	27	17	10	0

Comparison of Average increase (or decrease) in percentage mean marks with respect to the results produced from test 1 to tests 2,3 and 4 by groups of students.

GROUP	Average Test 2-1	Average Test 3-1	Average Test 4-1
10-19	14.25	22.95	11.25
20-29	16.13	10.50	22.50
30-39	9.19	13.44	19.59
40-49	3.34	6.90	13.46
50-59	-9.59	2.87	14.12
60-69	-3.00	-3.48	0.61
70-79	-12.75	-12.46	-2.18
80-89	-7.50	-8.64	-4.50

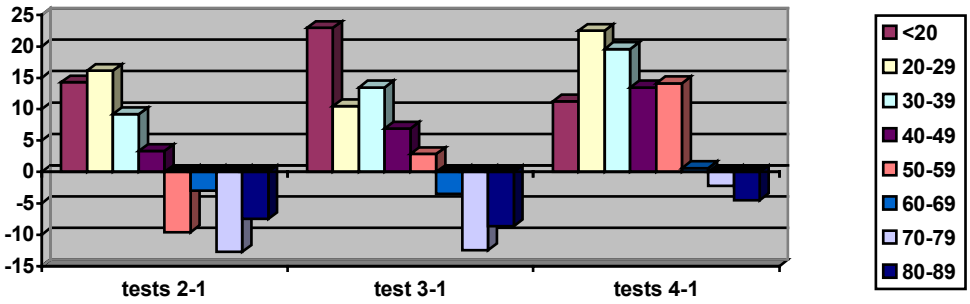

Figure 4

Analysis of Groupings

Figure 4 shows that the students who have achieved less than 49% on the first test, on average continued to show a positive improvement compared with their performance in their first test throughout the process of further testing. This indicates that these students initially had problems with the module, however improved throughout the course of the module, even though

the quantity and possibly quality of the work increased. This increase in performance of weaker students throughout the course of the module reflects well upon the module. Often in the past, due to the cumulative nature of the knowledge gain within a fairly technical module, the weaker students would get lost at the start of the module never to recover. This method of continual assessment indicates that a learning process has taken place throughout the module's progress.

Even though there was a drop in the overall average scores between test one and two (approximately 4.5%), the groups below 50% all improved, yet the groups above this 50% boundary all dropped in performance.

Students who achieved a mark of 70% or greater in the first test, on average failed to meet the standards of the first test.

The students who were to be initially found in the groups of 30%-39% and 40%-49%, showed a marked improvement throughout the course of all the tests.

Conclusions

The obvious conclusion from the introduction of this system, is the marked improvement in the results produced. The average mark and the number of students passing this module at the first attempt have improved dramatically. The students have been assessed making use of the **OLAL** system throughout the module's course. It has proved to be a very valuable assessment tool, which has produced instantaneous feedback to the students, and has made the marking process one of simple report generation. With respect to the learning process produced by the formative nature of the testing, without making any wild claims, the results appear to indicate that it has been a success and aided learning. From the student feedback this has been confirmed.

From the feedback, the students really enjoyed this method of assessment, which is a mark of its success. The peer pressure aspects provided a source of humour to the testing, with students enjoying the signalling of a colleague's incorrect answer by the beep. This never became a source of ridicule to any of the students as they were allowed to turn it off if they were concerned with this problem. Very few turned off the sound. In fact the OLAL system is "affectionately" known as "the BEEP test" within the school of computing.

The major argument against the claims of the "learning" aspects of the system, is obviously, that the students merely remember the position of the correct answer and then select it by memory rather than context. If this were true, then the results would not be expected to improve throughout the duration of subsequent tests as is shown in the analysis.

Question randomization, and the movement of the distracters, could limit this argument, and may be utilized in future versions of the **OLAL** system. Future work will also include the use of a third pass of the test, presented a week

later to assess if there is a significant learning factor in the testing, and also an analysis of the results of the first and second attempts at each question.

References

Carneson,Delpierre,Masters..... Designing and Managing MCQ's, University of Capetown

Dempster ..... 1994 "Question Mark Designer for Windows" in Darby and Martin(Editors), Active Learning No 2 CTISS Publications Oxford

Field & Chandler ..... 1993 "Portfolio Assessment" Maricopa Centre for Teaching and Learning

Gardner&Medwin ..... 1995 Confidence Assessment Association for Learning Technology Journal Vol 3

Haladyna T.M. .... 1994 Developing and Validating Multiple Choice Test Items Lawrence Erlbaum Assoc. Inc. Pages 61-86

Haladnya & Downing A taxonomy of multiple choice item writing rules Applied Measurement in Education 51-78 1989

Snow & Lohman 1989 ... Implications of cognitive psychology for educational measurement ... Educational Measurement 3rd edition 263-332