Correlation between academic and skills-based tests in computer networks

William Buchanan

Dr W.J. Buchanan is a Reader in the School of Computing at Napier University, Edinburgh. He has published over 20 academic books in computing and electronic engineering. At present he teaches networking and security, and his research relates to distributed systems and mobile networking. Over the past decade, he has published over 60 journal and conference papers. His WWW page is at http://buchananweb.co.uk. Address for correspondence: Dr W.J. Buchanan, School of Computing, Napier University, EH10 5DT, Scotland, UK; email: w.buchanan@napier.ac.uk

Abstract
Computing-related programmes and modules have many problems, especially related to large class sizes, large-scale plagiarism, module franchising, and an increased requirement from students for increased amounts of hands-on, practical work. This paper presents a practical computer networks module which uses a mixture of online examinations and a practical skills-based test to assess student performance. For widespread adoption of practical assessments, there must be a level of checking that the practical assessments are set at a level that examinations are set at. This paper shows that it is possible to set practical tests so that there can be a strong correlation between practical skills-based tests and examination-type assessments, but only if the practical assessment are set at a challenging level. This tends to go against the proposition that students who are good academically are not so good in a practice test, and vice versa. The paper shows results which bands students in A, B, C, and FAIL groups based on two online, multiple-choice tests, and then analyses the average time these students took to complete a practical online test. It shows that there is an increasing average time to complete the test for weaker students. Along with this, the paper shows that female students in the practical test outperform male students by a factor of 25%.

Introduction
Academic testing in computing subjects has generally moved from an exam-extensive assessments into assessments that are more based on courseworks. Unfortunately, the move to coursework-based testing has proven difficult in the computing discipline for many reasons, including:

- The large sizes of classes. Computing-related programmes often have class sizes measured in 100 s, which makes the marking of courseworks difficult, as it is often a
time-consuming task. For example, an academic will take over 40 hours to mark 250 courseworks (assuming that they spend 10 minutes reading each coursework). Thus, this time can thus often be as long as the total time spent on teaching the module.

- Large-scale plagiarism. Unfortunately, most academics in computing have seen extensive examples of plagiarism. In the past this was mainly focused on copying from fellow students, or small-scale copying from books. With the Internet, there has been a massive growth in verbatim copying from WWW pages. This is now a problem that is made worse with the large-scale variation of the rules relating to plagiarism across institutions, and even across academic departments within the same institution.

- Plagiarism due to weak coursework specifications. For courseworks a common complaint is that the coursework specification is not well defined and does not reward students for well-constructed research.

- Franchising of modules. Many universities have franchised modules for Higher National Diploma (HND) and degree-level modules. These can cause problems, especially as franchised colleges adjust to marking to the same level as their parent institutions, and acquire confidence in certain technical areas. In many cases the quality of a module at a franchised college can be based on the one-to-one interface between the module leader in the parent institution, and the module leader in the franchised college. If this academic relationship does not work, or if there is no regular communication, the module can hide problems that are not uncovered until it is too late. Problems can also occur when modules are franchised in another country. A good example of this is in modules run in Far East countries, which have different time zones, and typically run a semester system which is offset by many weeks as from a UK-based institution.

- Relevancy of exams/courseworks. Many computing-related students now demand practical skills relevant to the jobs market place, and may devalue academic skills, such as research skills, analysis and design, in favour of the practical skills, such as configuration and programming. Networking is a good example of this, as many students demand a good deal of hands-on experience, such as in the programming of routers and switches. The Cisco Academy (http://cisco.netacad.ne) has shown that there is a strong need for skills-based training in an academic environment, but it struggles to deliver the high level of academic rigour that is required at Levels 3 and 4 of degree programmes.

- Time factors in coursework completion. Many academics have observed that some students spend far too long in the creation of courseworks, and others spend very little time. It is, of course, difficult to define how much time a student should spend on a coursework. Deadlines have often been extended on coursework submissions as a result of students spending too much time on their reports (or through bad planning from the student). In a practical assessment students will know themselves when they reach the required level and, when ready, will take the assessment.

- Support time for courseworks. There is typically a large investment in time in both the support of a coursework and in its marking. The more challenging the coursework, typically, the more time that an academic will spend with an individual student.
discussing its specification and implementation. A practical test that can be practised in the student’s own time would obviously bring benefits in time saved.

An assessment model that allows these problems to be solved yet still achieves a good standard of academic rigour is obviously required. This paper outlines an example of a module that has been refined over the years and shows that it is possible to create a skills-based assessment that has a strong correlation with academic assessments.

In the teaching of the module presented in this paper, it has been observed that there is a degree of consistency in the performance of students where the grades that they get when they take tests tend to stay fairly constant. For example, an A-rated (over 70%) student will tend to get A’s or high B’s in examinations for similar modules, and a C-rated student will tend to get C’s in other similar modules. The link between examination grades and courseworks is also strong, as hardworking and well-motivated students tend to work hard in both examinations and in courseworks, whereas a weak and less motivated student will generally work less for both their courseworks and their exams. It is the objective of this paper to show that there can be a strong correlation between practical skills-based tests and examination-type assessments.

**Computer Networks module**

Over the past few years, the School of Computing at Napier University have run a successful Level 3 module in Computer Networks (http://www.dcs.napier.ac.uk/~bill/cnds.html). It is taken by over 250 students, of which 200 are Napier-based and 50 are from other UK-based institutions, along with many students taking the module in Malaysia. The module had been refined over the years, so that there was no final exam, as a final exam does not give student much chance to gauge their performance on the module before it ended. Along with this, courseworks were not seen as the most efficient method of assessing student skills, as many students simply used the Internet to obtain their material, especially in the theory and background sections of the coursework, or cribbed results from other students. Thus, it was decided that the module would be based on three assessments.

With such large numbers of students on the module, it was decided that the tests would be taken as online tests, which allowed more work to be put into their design, and less into their actual marking. The two online tests were carefully created with a defined mixture of questions that were graded for difficulty. These assessed the following academic skills:

- knowledge;
- mathematical reasoning; and
- research skills.

and covered the following learning outcomes:

L1. Outline the key elements of network infrastructures and networking components.
L2. Develop an understanding of network addressing, and how networks can be subnetted.
L3. Understand the fundamentals of network programming, and identify the key elements of client/server and peer-to-peer programming.

The last two learning outcomes were:

L4. Identify the key elements of router configuration and routing protocols.
L5. Identify the main elements and parameters involved in networks that have mobile devices.

which were assessed by using a skills-based test (http://www.dcs.napier.ac.uk/~bill/network_emulator2.htm). For the first set of assessments, it was decided that the first two tests would be graded to give an overall mark for the module (L1–L3), and that the skills-based part would be used to define a pass/try again assessment (L4, L5). With the results of this skills-based part, a future grading strategy could be created, and research could be conducted as to the relevancy of grading for the practical test.

**Academic testing**

All students took online examinations in weeks 6 and 10, and these were used to grade the students for the final mark. A template of questions was used for the assessment, with 35 questions with a defined specification, such as:

Q1. Identify the default subnet mask for a Class A/B/C network.
Q2. Determine the number of usable subnets for a Class A/B/C network and a given subnet mask.
Q3. Determine the number of usable hosts for a Class A/B/C network and a given subnet mask.

The questions that the student received will thus be of the same difficulty level, but each question differs in a key element, such as asking the student to calculate a bit rate for a defined number of transmitted bits over a given time. For example, using a template of:

```java
```

where A is the question number and n is the instance of the particular question. An example of a question with multiple definitions is:

```java
question15[0] = new myquests("A network uses a Class B address with a subnet mask of 255.255.248.0, determine the number of usable host on each subnet: ", "254", "510", "1022", "2046", "4094", "D");
question15[1] = new myquests("A network uses a Class B address with a subnet mask of 255.255.252.0, determine the number of usable host on each subnet: ", "254", "510", "1022", "2046", "4094", "C");
question15[2] = new myquests("A network uses a Class B address with a subnet mask of 255.255.254.0, determine the number of usable host on each subnet: ", "254", "510", "1022", "2046", "4094", "B");
```

and so on.
With the results of the tests, it was observed that students who did well in the first assessment also did well in the second one. The students were then graded by using: A+ (above 85%), A (76–84%), A− (70–76%), B+ (65–69%), C− (40–45%), and FAIL (less than 40%). The number of students in each of the bands is shown in Table 1.

The data show a good spread of marks. After each test, the students were given formal feedback from the tests, and it was observed that many students who had failed the first two tests had worked very little and had not asked for help with problems that they had. A common problem was that they had not read the additional material that they were directed towards. The early completion of the assessments which counted for the final mark gives students the opportunity to be reassessed before the end of the module.

As previously mentioned, it has been observed in the past that there was a strong link between good performances in one test. This can be observed from the results of the top 20 students (http://www.dcs.napier.ac.uk/~bill/cnds_student_test00.html) (see Table 2).

The results show that students gaining an A in the first test tended to gain an A in the second test. Students at the lower end also showed a level of consistency in gaining Cs in both tests. This type of result can thus be used by academics to quickly identify weak students. It can also be used to identify potential first class/upper second class Honours students, and potential lower second /third class Honours student.

Skills-based assessment
The third test was designed to a practical hands-on test, and taken in week 13 of a 15-week semester. With such large numbers on the module it would not be possible to allow each student to be assessed in a practical way, as Cisco Academy students are; thus, an emulator was created to assess practical skills. Figure 1 shows the interface that the students used. It produces a unique test with differing parameters to set every time when it is executed (this stops students from learning the material in a parrot fashion).
The student then configures three networking devices: a router, a switch, and a wireless access point. Examples of configuration include:

• setting the network addresses for all the ports of the devices;
• setting the routing protocol;
• setting device names;
• setting up virtual networks; and
• setting up wireless parameters.

The system then detects if the student has a correct answer, and after 25 correct answers, the system stops the online system, and registers a pass. The time taken to complete the test is then registered (see Figure 2). At present this is time is used for the research purposes on their performance, especially in the correlation between their performance in the first two online tests and in the practical assessment.

To reduce stress on the student, they are allowed to use the emulator for practice, and only take the test when they are ready to complete it. An instructor then enters a password, and when they complete the test, the time taken to complete the test is stored, along with their configuration steps. Students are informed, immediately, that they have passed the test, in order to receive quick feedback. Only the top times are published, so that high-performing students can feel proud of their achievement.

**Results**

The results from the first two tests were graded and assessed against the average time to complete the online skills-based test. The results are given in Table 1. Whilst there is...
variation within the grades for the test, the trend can be easily seen when the A, B, and C grades are taken as average values, as given in Table 3. It can be seen that there is now a strong correlation between the performance on the academic tests and the skills-based ones. It can also be seen that students who failed the first two tests took nearly 80% more time to complete the skills-based test, on average, than an A-rated student. The largest gap is observed between the B-rated student and the C-rated student, where there is a 45% increase in the average time. This, again, is an observable attribute, as B-rated students tend to understand many of the principles involved, but maybe do not work as hard, and are not as motivated, on average, as A-rated students. Typically, students who score Cs often struggle to understand the basic concepts of the subject and do not ask questions until it is too late (such as after the assessment).
Another observation is that the variation in times is the largest at the lower grades. For example, the FAIL, C–, C, and C+ grades have a standard deviation of around 600 seconds, whilst the A and B grades have a standard deviation of 240 seconds.

Another trend, which is seen from the practical test, is that female students did, on average, much better than male students, where the average time for students was:

Male students: 898 seconds  
Female students: 716 seconds

This can be contrasted with the average marks for male and female students for the first two tests which were:

Male students: 55.9%  
Female students: 55.6%

It can be seen that this is a similar mark. Thus female students outperformed male students by more than 25% in the skills-based test, which is an encouraging
result, especially in attracting female students to information technology/technical subjects.

**Conclusions**

Computing modules in Higher Education have many challenges, such as the growth in student numbers, and in the large-scale plagiarism from the Internet in courseworks. This has caused academic departments to search for an optimal way to assess students, especially in creating assessment strategies that allow more time to be spent in creating assessments and in teaching students, than in marking student scripts. The move from examinations to coursework-based assessments has not really helped this as it can take just as long to assess courseworks properly as it does to mark exam scripts. In contrast, the move to online examinations brings obvious benefits, as more time can be spent on designing the questions, and virtually no time is taken to mark students’ work. It also allows more diagnostic work to be completed so that weaker students can be quickly identified and given help. However, online tests must be carefully designed and should not fall into the trap of vendor-provided assessment strategies that tend to focus on factual knowledge as the main assessment criteria. If they are not designed correctly, students can learn the answers to the tests by gaining the standard answer, and spend very little time doing background reading. The model presented here uses a template of questions, which define each of the questions, each question differing in a key element, such as asking the student to calculate a bit rate for a defined number of transmitted bits over a given time. The number of bits and the time varies over the questions given to the student, but they are all at the same level of difficulty.

The model defined in this paper is helpful as it allows students to be graded at an early stage, where the rest of the module can be used to develop practical skills that students find useful in gaining employment. A key element is that the skills-based part is assessed in a formal way with some form of challenge to allow students to stay motivated throughout the module. It was observed that many students were continually scanning the top times for the skills-based test to see if they still had a top time. This generated positive competition where no student felt a fear of failing the assessment, but instead actively competed against others in the time taken.

The results also show that online practical tests can be made challenging so it is possible to create assessment strategies that have a strong correlation to student performance in academic tests. From these tests it can be seen that a benchmarking process can be used to grade students for the practical test and these can then be applied for future tests. Continual refinement is obviously required and the tests should be contrasted against traditional exam-based testing.

The use of emulators is beneficial in assessments and in allowing hands-on experience, but they cannot fully replace real hands-on with wires and boxes and, if possible, students should also be given some time on the configuration of real-life devices. Also, whilst determining the time taken to complete an activity is possibly not the most efficient way to grade a student in this area, it at least proves that it can be
used as a metric in the testing. At present work is being completed on an analysis in the flow of commands that a student enters, and marks are allocated to the most logical approach. Along with this a fault emulator has been developed which generates a number of unique faults (Figure 3) (http://www.dcs.napier.ac.uk/~bill/network_emulator1_fault.htm). The time taken to debug the faults will also be a good measure of the student’s performance on the practical side of networking. Fault-finding and debug/testing are two of the most important skills in the computing industry. Academia typically avoids any mention of them, especially in fault finding.

Whilst the model presented is not suitable for every subject, it seems to be useful in an area of computing which depends on both academic and practical study. Most academics could identify practical elements in their subject area that could be used for a practical test, such as in a programming exercise for the time taken to determine the fault in a computer program.

Overall, students seem to enjoy the challenge of the online practical test and a degree of positive competition was observed. Whether this positive competition would still be relevant if the students were graded for their performance is another area of research.