Science and Mathematics Education Centre

# Computer Managed Learning Assessment in Higher Education: The Effect of a Practice Test.

Janet Leith Sly

"This thesis is presented as part of the requirements

for the award of the Degree of Doctor of Philosophy

of the

Curtin University of Technology"

**October 2000**

# ABSTRACT

This thesis reports the results of studies set up to investigate formative assessment in the context of a computer managed learning (CML) practice test. The studies sought to determine whether taking the practice test affects performance on later CML assessed tests for first year university students and to determine the characteristics of the most effective CML practice test. The study was carried out in the context of CML testing at Curtin University of Technology. Because data were collected in a real testing environment, the research questions were addressed using a series of small studies, each focusing on a one-semester unit for first year students. Those students who sat a practice test improved their performance from the practice to the assessed test. Further they outperformed the non practice test group on the assessed test. The effect was statistically significant in eleven of the twelve studies where CML test results were investigated. Student ability, anxiety level, and sex did not affect test performance or choice to sit the practice test. Students preferred to be given the correct answer for an incorrect response and to have a practice test that was the same length as the assessed test but students continued to show improved performance regardless of these conditions. They reported that they used the feedback in a variety of ways including identifying important areas of content, identifying their own error areas and as a motivator for further study. The findings suggest that using the CML system as a formative assessment tool improves student performance on summative assessment. The practice test is contributing to improved performance, however this improvement cannot be attributed to a single factor. In those cases where the practice test only partially covers the content of the assessed test, the improvement is seen on that common part, however when there was no overlap of content the group who did the practice test still performed better on the assessed test than the group who did not. This suggests that a contributing factor may be familiarity either with the CML system, items or test type. It is also possible that the beneficial effect was due to prior exposure to the CML system and that only one test is required for this purpose.

This research has implications for current teaching practices because the acceptance of a practice test provides feedback to both students and lecturers prior to the assessed test. The optimal practice test covers the same content as the assessed test with the same number of items and provides the correct answer for a item answered incorrectly. The key recommendation for use of the CML system is the provision of a practice test for formative purposes, for the use of both lecturers and students. Lecturers need to encourage student participation not just on an initial practice test but on all practice tests provided. Students need to be encouraged to review their error summary, as is the current practice in the CML Laboratory. Lecturers need to make more use of the feedback provided by the tests, in terms of content coverage, revision and consolidation of work, and quality of test items.

# ACKNOWLEDGEMENTS

There are three people without whose help this thesis would not have reached this stage. I wish to thank my supervisor, Professor Léonie Rennie for her valuable advice, her untiring patience and her attention to detail, my friend and colleague, Jenny Lalor for her expert and always patient help with the data analysis in this thesis and my husband, Peter, for his support, encouragement and total understanding at all times.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION TO THE STUDY

There is a broad range of assessment methods presently in use across educational institutions. Included, among others, are written examinations, standardised tests, practical sessions, teacher assessment, peer assessment and computer-based testing. While methods vary the key role of assessment should be to help the individual develop and learn. Assessment itself should be a powerful tool both for developing learning and increasing student motivation. Gipps (1994) captures this sentiment: "What we wish a good assessment to do is to elicit quality performance, that is the pupils' best performance, within a well defined context" (p. 290).

Assessment occupies a central place in the life of any undergraduate and it is becoming increasingly common for this assessment to be computer-based, especially with first year undergraduate students. The reasons are varied but include the need to accommodate large classes with less available lecture hours and resources as well as the need to manage progressive testing. Computer-based assessment when used for summative purposes affects students' final marks and so, following Gipps' sentiment, the aim of this research is to identify the most effective parameters for a computer-based formative test that will "elicit quality performance" from students on their summative test. The context is well defined. It is computer testing at Curtin University of Technology for first year undergraduate students. Testing is performed with a computer managed learning (CML) system used through a central testing laboratory.

As CML systems are used more widely to generate student assessments, there is a need to provide students with the best opportunity to perform to their potential rather than accepting that present usage, a single summative test, is adequate. At the start of this research, in 1997, only one unit using CML at Curtin University of

Technology offered students the option of any formative assessment using CML testing. While twenty units are tested through the CML Laboratory each semester, with approximately 11,000 students doing up to 30,000 tests in any one year, only a few of these units use a practice test. Prior to 1998, Economics was the only unit with a practice test and this was before the first assessed test only. When lecturers use a practice test many often believe its value is only to familiarise the student with the working of the computer software package. As the results of early studies became known more units began using the practice test facility, however, it was only in first semester 1999 that any unit ran with a practice test prior to both assessed tests. Less than a quarter of the units using the CML system presently use the practice test facility.

This research has been instrumental in increasing this use. Most CML systems have the ability to code tests and in this study this functionality is used to identify some tests as practice tests. The formative aspects of these CML-generated multiple-choice tests designated by the lecturer as practice tests are investigated. The general research question asks whether the practice test affects later CML performance and what characteristics make it most effective.

## Background and Context

### Computer Managed Learning Systems

Computer managed learning systems are software packages with several common features, including generating tests from banks of items, marking of tests generated, analysing of the results and keeping records of students' marks and progress. Feedback to students on their responses is common. CML systems usually hold many banks of items, each of which can be categorised into smaller components, such as topics, learning outcomes or objectives. Often items can be coded on a variety of characteristics, for example, degree of difficulty and cognitive content. So, we have test banks, which can be subdivided into component parts, made up of items which themselves can be categorised.

The key function of a CML system is to generate tests. It has a template or a course-map where the parameters used to generate any test are stored. There is a range of available options depending on the specific CML system in use. Tests may be generated using the same items for all students, others may be set to select items randomly but to take them from specific parts of the bank of items. In this case, students are more likely to have unique tests, depending on the number of items available for selection. Another option could include the coding of individual items as a variable, for example, a test could be set to draw a specified number of items from selected objectives but to take only those items coded at a particular difficulty level. Again, the test could specify that certain items were mandatory.

Most systems are able to generate tests that use multiple-choice, true-false, short answer, matching and calculation items. Most also have the ability to give students immediate feedback on incorrect answers and supply them with the correct response. Some give students additional information in the form of comments on their response. Because systems retain students' answers to items, they are able to produce a report of the distribution of item responses and students' marks. This feedback can alert lecturers to problem areas in terms of item quality and also student performance. Because CML systems retain information about incorrect, as well as correct, answers, lecturers can be made aware of common misconceptions held by students and be able to address these before summative assessment takes place.

The term Computer Managed Learning is used to describe the process by which a computer issues tests and manages student progress through a unit of study. The computer generates and marks tests providing a feedback function rather than a direct instructional role. Computer Based Education (CBE) encompasses any tutoring or testing of students which is performed with computer assistance and many of these systems have a CML component, however, they are not strictly CML if they perform a direct teaching role. Computer Based Instruction (CBI), Computer Aided Instruction (CAI) and Computer Assisted Learning (CAL) are all synonymous

terms for the portion of CBE which involves teaching the student. CML systems are presently being run from a variety of different platforms and some of these are described in the next section.

Mainframe-based CML Systems

When the system is mainframe-based, the CML software is held on a mainframe computer and the individual user accesses this system either from a dumb terminal or a personal computer (PC) using a terminal emulation package. To date there have not been any client/server systems developed. The CML software resides in one place but can be accessed from multiple locations. A key advantage of this set up is that when changes are required to either items or parameters for test generation they are performed once on the mainframe and are then available to all users. One example of a mainframe-based system is the Learning Management System (LMS) developed by Campus America. This is the system in use at Curtin University of Technology. According to its designers (Computer Based Training Systems, 1995), "LMS is designed as an instructional development, delivery, and evaluation tool to assist in the management of learning. It does not replace traditional instructional vehicles, but rather enhances and complements these techniques through their effective administration" (CBTS, 1995, p. 1-1). Its main functions are testing, evaluation, and control of student progression through a course of study. It is a menu-oriented CML software package designed for, and implemented only upon, the VAX/VMS platform.

Personal computer (PC) based CML Systems

Personal computer-based CML systems have the ability to operate in a number of ways. Options include the software residing on a central machine and being stored on each individual PC. When a central server-based installation is in operation it is similar to the mainframe system in that the software is shared by many users. However, it has been argued that the security of a mainframe-based system is superior, given the maturity of the operating systems. Another option is for the

personal computer-based CML software to be installed on one stand-alone machine. This setup cannot easily accommodate the testing of large student numbers and can create a bottleneck for processing. There are many personal computer-based packages available some of which are totally CML and others have computer testing and student management as a component of a bigger package. In situations where CML is one component of a larger package it typically has less functionality than packages that are totally CML with no instructional component. Examples of personal computer-based software packages include, among others, Elmsoft, Jarvis Education Package, Manager, and Tutor.

Web-based CML Systems

The advent of the World Wide Web now allows the delivery of course materials on a truly platform-independent medium. Often software packages that function in this domain have the ability to generate tests, provide feedback and manage student progress so can be included in the CML category, for at least part of the package. The new variable that this form of delivery introduces is the ability to provide access from almost anywhere in the world. However, the price paid for this total flexibility is the almost complete loss of direct supervision of the student who is doing the test. While, at this stage, these web-based package are very good as formative assessment tools, they are not able to support summative assessment because their flexible design lacks the ability to ensure testbank security or test security i.e., provide positive identification of the individual student taking the test. While password protection does allow student identification, by the very nature of the system, access can be from anywhere and while the individual may have identified themselves to the system with identification number (ID) and password, they may have helpers (books, tutors etc.) while they do the test. As an overall package these systems have many benefits for the student, including improved opportunities to interact with the actual learning materials, improved opportunities to interact with other people (email) and the ability to support graphics in a much more sophisticated manner than some of the older CML systems. Examples of products in this web-based category include, among others, WebCT, TopClass and WebMCQ.

## Computer Managed Learning at Curtin University of Technology

Curtin University of Technology uses a mainframe-based computer managed learning (CML) system to generate and mark large numbers of student tests each year. These tests are predominately multiple-choice and used for summative assessment. Items are drawn at random from a test bank according to parameters requested by the lecturer. The component parts of the testbank are modules, which are subdivided into objectives. Each module must have at least one objective but can have up to nine objectives. Items are stored at the objective level. While the system handles all item types mentioned earlier, most lecturers select multiple-choice. Although advantages may exist for using items which require written answers, i.e., constructed response items (Birenbaum & Tatsuoka, 1987), increasing student numbers and lecturer workloads have been associated with the use of multiple-choice items which are marked by the system. Students access both practice tests and assessed tests, but only the assessed tests contribute to the overall grade. The typical contribution of CML tests to the overall assessment is twenty percent.

CML systems for assessment and student tracking have advantages in managing large numbers of student assessments. If lecturers are to use CML as one part of their assessment program, it is important that they use effectively all of the information available. The tests generated using the CML system provide feedback relating to both student performance and the effectiveness of the items in the question bank, hence the information can be used for evaluation of both the students and the items.

## Formative and Summative Assessment

Scriven (1967) first used the terms formative and summative in the context of curriculum evaluation, but they now have wide currency in assessment. However, formative assessment does not have a strictly defined meaning. Black and Wiliam (1998a), in their review of the literature on classroom formative assessment, interpret it as "encompassing all those activities undertaken by teachers, and/or by

their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (p. 7). Sadler (1998) refers to formative assessment "that is specifically intended to provide feedback on performance to improve and accelerate learning" (p. 77). Rolfe and McPherson (1995) see it as a testing method without academic penalty that reveals short comings in students' understanding while allowing them to take responsibility for their own learning. In all cases the aim is to improve a process at a stage where change can be made rather than to appraise a finished product. In contrast, summative assessment is generally an end point assessment that contributes to students' grades. The terms formative and summative refer to the purpose of the assessment rather than to the methods used (Brown & Knight, 1994; Rowntree, 1987), thus the same test may be regarded as formative when it provides only feedback on performance, but summative when the mark contributes to the final grade.

The Contribution of Computer Managed Learning to Formative Assessment

As the type of assessment used at university level has the potential to influence student grades (Brown & Knight, 1994; Heywood, 1977; Nightingale, Wiata, Toohey, Ryan, Hughes & Magin, 1996; Piper, Nulty & O'Grady, 1996; Webster, 1992) the contribution of formative testing to improve student performance will be discussed in the context of CML testing. The features of the CML system used for formative assessment are the possibility of a practice test, the feedback to students and feedback to lecturers. Some lecturers allow students to do a practice test that is computer marked, and although the results are available to both lecturers and students, these scores do not contribute to students' grades. Importantly, feedback is available to the student on incorrect responses and the student is encouraged to review their test paper and error summary under supervised conditions in the CML laboratory. Often lecturers provide a textbook that can be used for reference in this review process.

Two kinds of feedback are provided to lecturers. They receive a statistical analysis of the responses to items and also a performance report, which lists all students enrolled in their subject with their test marks. The performance report alerts lecturers to students who are falling behind, and the statistical analysis is useful in identifying those content areas in which students are experiencing difficulty.

## Research Questions

The general research question asks whether the practice test affects later CML performance and what characteristics make it most effective. Specifically, the study will investigate the following questions:

1. How does the practice test affect performance on later CML assessed tests for first year university students?

   1(a) What is the effect of the amount of content overlap between the practice test and the assessed test?

   1(b) What is the effect of the length (number of items asked) of the practice test and what length do students prefer?

   1(c) What is the effect of feedback on incorrect responses and how do students use the feedback?

2. What are the characteristics of the most effective CML practice test?

## Method

The research used a multiple study approach involving 14 groups of first year undergraduate students from six subject disciplines, Economics, Psychology, Accounting, Human Biology, Education and Medical Sciences. All students were

enrolled in one-semester units and their marks for tests were obtained directly from the CML system for 12 of the groups. The other 2 groups participated by way of survey or interview only.

Research Question 1 was addressed across the 12 studies that used a practice test. Each of the component parts, 1(a), 1(b) and 1(c), of this research question were investigated in several studies. Question 1(a) concerned with content was addressed by comparing the results of those studies where the practice test covered the same content as the assessed test and those where the practice test covered less content than was examined on the assessed test. Question 1(b) regarding test length was addressed by those studies where the practice test and the assessed test covered the same content range but used a different number of items. The feedback issue in question 1(c) was investigated by three studies in which the practice test informed students merely that that an answer was incorrect but did not give the correct response (described as limited feedback). Additional information on students' desired practice test length, students' use of feedback and preference for mode of feedback was obtained by survey and interview.

Additional factors investigated as having the potential to influence results were students' subject discipline, familiarity with the CML system, ability, anxiety, sex and response to CML. Response to CML included students' general attitude to CML, their use of the practice test and their reported use of the feedback provided by the CML system. Subject discipline as a possible confounding factor was considered across a compilation of all studies. Familiarity with CML was investigated in two ways: firstly, by considering a group of students who had previously used CML for their prerequisite subject in the same discipline and, secondly, by considering a group whose practice test did not reflect the content of the assessed test, i.e., 0% overlap. Three performance measures that reflected student ability were investigated. They were university entrance score, final examination mark for the unit and final examination mark for the prerequisite unit. Each of these performance measures was investigated for its effect on students' choice to do a practice test in a

different subject discipline. Students' anxiety level was investigated by a self-report questionnaire used with two groups of students from different subject disciplines. Student surveys were used to collect information regarding students' sex and general attitude to CML while interviews were conducted to obtain additional information regarding the feedback facility provided by the CML system and to validate responses given on the surveys.

## Significance

The study is significant for three reasons. Firstly, it provides information on the use of formative assessment in the computer managed learning environment that has not previously been reported. It will lead to a more informed understanding of the use of CML as a tool for learning rather than as a summative testing technique. Secondly, the study can contribute to improved understanding of students' use of feedback, and the feedback provided by the CML system has the potential to enhance students' control of their own learning. Finally, it has implications for current teaching practices, because feedback about the practice test results which is given to lecturers prior to summative assessment, can be used to influence the pace of progression through the unit and the time spent on content areas that students find difficult.

## Limitations

The research design was constrained as all testing took place in an operational CML testing laboratory at Curtin University of Technology. Consequently, the research was opportunistic, making use of CML as it was used in the normal operating environment. Because of the pressures of academic work, it was not possible to structure the data collection through a series of purposely designed studies, rather data collection occurred using the testing format dictated by the lecturer. Further, it was not appropriate to manipulate the testing environment for some students and not others. Hence, data were collected for those groups where the testing situation was suitable to the research and where lecturers gave permission and students were agreeable to data being used. In this way, a picture of the CML system and the

effects of the practice test could be built up using a series of studies. However, it is unlikely the findings can be generalised to systems which do not operate in a very similar way to the one used in this study.

## Overview of the Thesis

This introductory chapter has outlined the major elements of the thesis, its background and rationale, purpose and significance, as well as giving a description of the specific context of the study, first year undergraduate CML tests on a university campus in Western Australia. The literature review in Chapter 2 includes the areas of assessment with specific reference to CML systems and the types of tests generated as well as the different functions for which these tests can be used. It reviews both student and test attributes that affect performance. The research design and method for the studies are described in Chapter 3. Chapter 4 presents the quantitative test results obtained directly from the CML system as well as that qualitative data that specifically addresses the first research question. In Chapter 5 the survey and questionnaire results from across the various studies are discussed as well as the interviews with students. This chapter specifically addresses those additional factors that have been identified as having potential to influence test performance. Finally, Chapter 6 provides a summary of the whole study, a synthesis of its outcomes and suggestions for future practice.

# CHAPTER 2

# LITERATURE REVIEW

This chapter introduces the important issues of assessment and assessment techniques. The various computer generated assessment techniques available are described, leading to a discussion of formative assessment and its importance to learning. In the following sections, feedback and self-regulated learning are defined. The functions of feedback and factors affecting feedback reception are then placed in the context of both formative assessment and self-regulated learning. The chapter next addresses student attributes and the CML practice test, both of which have the potential to affect student's performance. The chapter concludes with a summary of the literature and implications for the present research, which will focus on the use of the computer to generate formative assessment in the context of CML testing at Curtin University of Technology.

## Assessment and Assessment Techniques

All undergraduates spend a considerable amount of time engaged in some form of assessment. Organising assessment also takes a large proportion of the lecturer's time and resources and impacts on students in ways that may be positive or negative. Crooks (1988) classifies the effects of assessment on students into groups based on a time perspective. He considers effects at the level of a particular topic or assignment (short term), a course or extended learning experience (medium term), or the longer term consequences that occur when students meet consistent patterns of assessment from one year to the next (long term). At the short term level, Crooks mentions eight key effects, including the consolidation of prerequisite skills, focusing attention on important aspects of the subject, encouraging active learning strategies, providing corrective feedback and helping students both to monitor their own progress and to feel a sense of accomplishment. While these are important consequences, as the levels progress the effects become more critical. At the course or medium term level,

Crooks refers to the certifying of students' achievements as "influencing their future activities" (p. 443), while longer term effects include influencing students' motivation, development of learning skills and styles as well as their self perception.

Depending on the contexts in which the results of assessment are given, and the ways students receive them, the effects can be positive or negative. For example, students might respond positively to assessment, working more effectively. Others may become disheartened and lose motivation. As the effects of assessment potentially have such critical consequences for students, it is essential that all assessment is well planned and tailored for its particular purpose. Another important aspect is that for any given assessment technique there will be some students who perform well and others who do not. Partly for this reason, Brown and Knight (1994) argue that the use of multiple techniques is essential to good assessment practice, because this does not disadvantage any one group of learners.

When selecting a technique for a particular assessment, an additional issue that needs attention is educational relevance, which Rowntree (1987) views as the key determiner of the choice of assessment technique. Assessment techniques need to be appropriate to the content and style of learning, i.e., essay examinations would not be appropriate for a course that had focused on laboratory practicals, nor would a multiple-choice test be appropriate for an oral language course.

There are many kinds of assessment techniques, besides the traditional final examination favoured in many university courses. Other assessment techniques used include multiple-choice or short answer tests, essays, assignments, bibliographies, oral seminars, written laboratory reports, and portfolios of student work over time. Assessments may be computer, peer or self-marked and in some cases they may be computer-generated. This study deals with a subset of assessment techniques, namely, tests which are generated and marked by a computer. The following sections discuss those assessment techniques available for use with a computer-generated test.

## Computer-Generated Tests

Computer-generated tests can take a variety of formats ranging from those where the student does the whole test at the computer to those where the computer is used only to generate a paper copy of the test. The type of item used on the test varies depending on the software package and also on whether they are to be computer marked. While many banks of items are predominately multiple-choice (Donnelly, 1989; Harrison & Stanford, 1989) a major consideration is whether the test itself is a good test. "A good test is one whose results accurately measure what students know about a topic at the appropriate cognitive levels" (Trollip, 1992, p. 7). The test itself must be well planned and the items well written.

### Multiple-choice Items

Multiple-choice tests, computer-generated or otherwise, have a number of obvious advantages, especially as part of an overall assessment strategy. Brown and Knight (1994) cite the range of content which can be covered by the items, the ease of marking and the ease of identification of major content areas that students find difficult, as key advantages for multiple-choice tests.

Common criticisms often cited with reference to multiple-choice items are that they test only factual recall and are subject to guessing (Ebel & Frisbie, 1986; Morgenstern & Renner, 1984; Pritchett, 1999). In contrast, other researchers (Hill, 1981; Race, 1998; Trollip, 1992) regard multiple-choice items as versatile, arguing that they can be used to measure different kinds of content and levels of cognitive behaviour, including the higher levels such as application, analysis, and evaluation. If properly designed a multiple-choice test can identify and test a range of skills as well as reducing the possibly of success by chance (Gronlund, 1982).

Trollip (1992) suggests that the way to use multiple-choice items to measure different content levels is to plan tests thoroughly. He proposes a method of planning for tests that is based on a matrix. One axis of the matrix has a breakdown of the

curriculum being tested and the other has a list of categories from Bloom's taxonomy (Bloom, 1956). Each cell is then considered in terms of its relevance to the particular test with number of items and marks assigned accordingly. Trollip suggests that this is a method of ensuring that all content areas are considered at the appropriate cognitive level. However, like other forms of assessment, they should be related both in weight and content to the goals of the subject in which they are located (Isaacs, 1994). This is consistent with Rowntree's (1987) call for educational relevance.

Multiple-choice items are less sensitive to guessing than true-false items (Pritchett, 1999), although more sensitive than those items where the learner must create an answer, that is, construct a response, rather than simply choose a response (Warner & Thissen, 1993).

In terms of marking, multiple-choice items are generally considered to be fair by students because marks are unaffected by external factors such as staff expectations of particular students as being a capable, hardworking, deserving of a good mark, etc. Personality factors are not able to influence the grading process (Charman, 1999). Lukhele, Thissen and Wainer (1994) report research in which the essay test yields a lower reliability than a parallel multiple-choice test, when administration times are comparable. This is attributed in part to the subjective scoring of the essay test.

Race (1998) suggests that multiple-choice items test how fast students think, rather than how fast they write. He sees this as an added advantage of the use of this item type as there are students who are disadvantaged because they are unable to write down the information they know in the allocated time.

One major disadvantage of multiple-choice items is the considerable time required to write good items, especially those that test higher levels of knowledge (Trollip,

1992). There is also the requirement for a large number of items in an item bank if students are to receive different items from one semester to the next.

Multiple-choice items when used in the CML testing environment can be marked quickly and reliably and produce immediate feedback. Items can be selected randomly from a pool of items, or alternatively, the selection of the same items for all students can be randomised on a particular test so that all students have the same items but receive them in a different order. This latter option has the additional benefit that the pool of items does not have to be as large as would otherwise be needed. The disadvantages for multiple-choice items in the CML context are similar to those in any other assessment situation. Items take time to create and require expertise to be written to address the higher order cognitive skills. However, they can be a valuable part of computer generated tests if they are carefully designed and used with restraint as one part of an overall assessment plan (Cann & Pawley, 1999).

Constructed Response Items

Constructed response items include completion items, brief essay items and extended essay items. They are often described, rather inaccurately, as subjective items, because scoring requires expert judgment. While this type of assessment task is not fully supported by CML, as it is not able to be computer marked, it is mentioned here because it has features that multiple-choice items do not.

Constructed response items differ from multiple-choice items in that they require the learner to create rather than choose a correct response. Hill (1981) sees problems with scoring reliability as well as the time required per item for the student to construct an answer as disadvantages with this assessment type. As items are often marked by one or more lecturers, factors such as handwriting, neatness, etc., have the potential to influence the marker's judgment of a particular response. The time factor inhibits good sampling of content on a particular test.

Constructed response items are expensive to mark because they require a subject expert to read the responses. Warner and Thissen (1993) found that in the Advanced Placement testing program in the United States a constructed response test of equivalent reliability to a multiple-choice test takes from 4 to 40 times as long to administer and is much more expensive to score. Advantages of constructed response items include that they are easier to write in ways that test higher-level cognitive objectives, they are impervious to guessing and they allow greater insight into the thinking behind student responses.

One of the primary reasons for using the constructed response items stems from the perception that constructed response items can measure traits that multiple-choice items cannot. However, Barnett-Foster and Nagy (1996) reported an analysis of response strategies employed by 300 first year undergraduate students on a chemistry test. Items were in both multiple-choice and constructed response format. Analysis of the test data across the different formats showed no significant differences in the types of solution strategies students used, in the type of error committed or in the success rate of students. Their findings support the hypothesis of Traub (1992) that items that require application of learned information appear to be impervious to format effects. They did, however, note that there was a high incidence of the use of algorithmic strategies and a low incidence of proportional, propositional and higher order reasoning strategies for both item types.

Bennett, Rock and Wang (1991) also found little evidence for the stereotyped view of multiple-choice and constructed, or free response, formats measuring substantially different constructs, i.e., trivial recognition versus higher order process. Their conclusions were limited to the computer science domain, the College Board's Advanced Placement Computer Science (APCS) examination and the tasks presented. In these examinations, the free response items were developed to measure certain content more deeply than the multiple-choice items, however, Bennett et al. speculate that the underlying processes measured by these free-response items are also measured by the multiple-choice items. While no single multiple-choice item

will measure more than one of the processes, they suggest that a number of items might cover many of the processes examined by the free response items.

Tests for Advanced Placement Chemistry and Advanced Placement United States History were examined by Lukhele, Thissen and Wainer (1994). They concluded that the constructed response items provide "less information in more time at greater cost than do multiple-choice items" (p. 245). They also found no evidence to suggest that multiple-choice and constructed response items measured different things.

Perkins (1999) suggests that computer-based methods of assessment have the potential to be more reliable than traditional methods. He defines reliability in terms of consistency of scores across markers or over time. There is also some debate with respect to the reliability of the multiple-choice item type when compared with the constructed response type. Some researchers (Ebel & Frisbie, 1986; Green, 1981) report that multiple-choice items are more reliable than constructed response items, but others disagree (Aiken, 1987; Traub & MacRury, 1990, as quoted by Barnett-Foster & Nagy, 1996).

The type of study habits promoted by each item type is also not agreed upon. Adams (1964) argues that the constructed response item type promotes better study habits, while Sax and Collet (1968) suggest that differences may be small and may actually favour multiple-choice examinations. However, Biggs (1996) states that when students study for examinations they do so in ways that they perceive will meet assessment requirements.

The choice of item type in this research has been defined by the CML system and lecturer's choice. However, while tests are composed of multiple-choice items the test itself does not always have the same purpose.

## Purposes of Assessment

While tests can be identical in format they may be used for quite different purposes. The two main purposes of assessment are summative and formative. Summative assessment produces a measure that sums up someone's achievement and so its real use is as a description of what has been achieved (Brown & Knight, 1994). It is generally an end point assessment that contributes to students' grades and it may or may not include feedback to the student. Typical of this form of assessment is the final examination where students receive a final grade but no other feedback. A large proportion of lecturers at Curtin University of Technology who use CML tests use them solely for summative purposes. Students are allowed one attempt at the assessed test and their score counts towards final grades. They are informed whether a particular response is correct or incorrect and in most cases are given the correct answer. However, as students have no option to repeat the test and, in most cases, the material is not assessed again, it is unlikely that many students make use of the feedback.

In contrast, formative assessment is designed to gain an ongoing overview of student progress (Cartwright, 1997). Feedback is a key element because it provides information that the student can use to focus on areas of weakness. Coursework assessments where the student receives feedback, such as from diagnostic tests and practice tests which are used to help the student diagnose areas of weakness and improve performance, are examples of this type of assessment.

Feedback is consistently part of any definition of formative assessment. Black and Wiliam (1998a) and Sadler (1998) specifically refer to the role of feedback in modifying or improving the learning process, and Rolfe and McPherson (1995) see it as allowing students to take responsibility for their own learning. Formative assessment and the feedback it provides give students both the opportunities to assess and change their present skill or knowledge as well as the ability to improve a process before a finished product is reached.

For formative assessment to be effective, learners need to perceive a gap between their present position and their desired position of skill or knowledge and take some action in an attempt to close the gap (Ramaprasad, 1983; Sadler, 1989). As personal factors and beliefs about learning can affect both reception of the feedback and the decision on how to respond, the effectiveness of formative assessment will depend not only on the context of the feedback and the associated learning opportunities but also on assumptions about student motivation and self perceptions (Black & Wiliam, 1998a). While all assessment techniques potentially can be used for formative assessment this research considers only those tests that are computer generated.

Computer-Generated Tests Used for Formative or Summative Assessment

A small number of researchers (Boyle, Bryon & Paul, 1997; Charman & Elmes, 1998; Fleming, 1987; Stanford & Cook, 1987; Zakrzewski & Bull, 1998) have claimed to use computer-based testing as a formative assessment tool. However, not all use the computer-based testing in a manner that would completely fit the definition of formative assessment outlined above. Stanford and Cook (1987) and Charman and Elmes (1998) appear to have used the results of what they called formative assessments to contribute to final student grades. While these tests could still have had a formative role, if the student used the information supplied to take some action to remedy mistakes and improve their grasp of the subject, such tests would not normally be regarded as formative assessment when the marks contribute only to summative assessment.

Stanford and Cook (1987) investigated the effectiveness of computer-based assessment in an introductory economics subject that was taught throughout the year, i.e., over the two semesters. They examined evidence obtained from a student questionnaire and also the results from the end of year examination. The final subject mark was arrived at by a contribution from various forms of assessment, including class quizzes, two final examination papers, and a CML test assignment. The second of the final examination papers was divided into Part A and Part B. The marks for Part B included a possible 10 marks for CML tests. These CML tests were

on an "all or nothing" basis with students receiving 10 marks towards their Part B score if they completed the CML tests at the defined mastery level. Part B was out of 35 marks and scores were scaled to be out of 25. CML tests completed at the appropriate level (3 attempts were allowed) then added 10 marks to the scaled mark to give a mark out of 35. Hence, the CML tests defined as formative contributed to the final grade for the subject, though the three attempts did provide students with the ability to use the tests as a diagnostic tool to improve future performance. Stanford and Cook compared the unadjusted and unscaled marks on Part A and Part B of the final examination and reported that the results in Part B were statistically different from those in Part A. No statistics were reported. They interpreted the results as consistent with the view that student performance in Part B of the course had been improved through the implementation of CML. By comparing final exam results for 1986, the year prior to the introduction of CML, with those in 1987, when CML was used, they believed that student performance had improved by "as much as 40 per cent on the final examination in 1986" (p. 377). On the basis of their report, it is difficult to determine how Stanford and Cook arrived at this conclusion.

An additional component of this study by Stanford and Cook (1987) included a student questionnaire. Students were asked 11 questions including questions referring to their opinion on the use of CML for the entire course or with other courses, as well as whether the department should proceed with the system. Ninety one percent of respondents said "yes" to having CML for the entire course, 87% of respondents would like to see it in other courses and 96% of respondents wanted the department to proceed with the system. The conclusion Stanford and Cook reached was that students were positive toward the CML system used.

Charman and Elmes (1998) evaluated the introduction of computer-based testing into a first year statistics module in geography. They reported improved student performance on the summative examination by comparing examination results for students in the year before (1996) and after (1997) the introduction of computer-based assessments. While the researchers attributed the improved student

performance to the introduction of computer-based testing there is another possible explanation. Prior to the introduction of computer-based testing, a portfolio of 10 practical assignments was used, so it is possible that the replacement of portfolios produced the difference in student performance rather than the computer-based testing. The computer-based tests that Charman and Elmes refer to as formative do appear to contribute to the final student grade although they were not included in the calculation of results for the summative assessment comparison.

Boyle et al. (1997) investigated a small group of students across a palaeontology module containing five topics. They investigated a computer-based learning delivery method as well as computer-based testing in order to enable a comparison to be drawn between computer-based and traditional methods of teaching and assessment. Three of the topics were delivered by conventional lecture, two used computer-based learning and one of these two used computer-based testing. Mean scores between topics were compared and it was reported that computer-based testing was as effective as conventional testing. Charman (1999) interpreted the results of Boyle et al. as showing that the variability in student performance was related more to the topic than the delivery method.

Zakrzewski and Bull (1998) report on the implementation of a university-wide computer-based assessment system that delivers summative, formative and self-assessment to students. This mass implementation at the University of Luton has seen a dramatic increase in the number of students using the system, which uses "Question Mark Designer" software. The system, which was originally piloted in 1994, had approximately 9000 students sitting summative assessments and 1000 students undertaking formative assessment in the 1996-1997 academic year. Zakrzewski and Bull report the conversion of four biology modules, originally examined using only summative assessment, from optically marked multiple-choice format to the "Question Mark" format. Three of the four modules supplied students with computerised formative assessment. In two of the modules using formative assessment a 1.6 grade point average increase was reported while no change was

reported for the module that did not use formative assessment. The fourth module actually showed a decrease in grade point average which the researchers attributed to a substantial change in the teaching and learning strategies for this particular module from the previous year.

While there are a large number of articles in the literature on the introduction of computer-based assessment into assessment programs, Charman (1999) states that none have effectively evaluated the contribution of computer-based assessment to formative assessment. He argues that the only way to assess the impact of computer-based testing as a formative assessment tool is to compare summative assessment before and after the introduction of formative assessment. However, as the CML tests used for formative assessment in this research are optional, it may be possible to judge the impact of formative assessment by comparing student performance on the subsequent assessed test for those student who used formative assessment and those who did not.

Evaluation of the Effectiveness of Computer-Based Assessment

An additional issue relevant to computer-based assessment is the evaluation method used to judge its effectiveness. Almost exclusively the focus has been on qualitative assessment or on student opinion (Boyle et al., 1997; Burden, 1992; Burke, Critchley & Lessells, 1989; Charman & Elmes, 1998; Coomb & Cooper, 1992; Harrison & Stanford, 1989; Mansen & Haak, 1996; Parsons & Hunter, 1987; Sambell, Sambell & Sexton, 1999; Zakrzewski & Bull, 1998).

Burke et al. (1989) reported a pilot project to investigate the potential for using CML at the University of Limerick. Students used the CML system to generate nine exercise sheets during the term. These sheets were provided solely for the students' own benefit and an adviser was available on a daily basis to answer student queries and system problems. A questionnaire was distributed to the first year engineering class to test students' reactions to the CML system. Approximately 60% of the 164

students returned completed questionnaires. Overall students liked using the system and considered it an added benefit to their study. Although 43% of respondents had no previous computer experience, this was not regarded as a drawback to using the CML system as it was considered user-friendly. Fifty eight percent of students liked using the system and 9% disliked it. When asked whether they thought CML contributed to their understanding of the subject, 66% responded positively and 14% did not think CML contributed to their understanding. Seventy seven percent of respondents wanted the release of correct answers with each exercise. Students reported that the main advantages of the CML system were that it forced them to work harder, was self-paced and was available all the time. The main disadvantages were the slowness of the printer and the limited feedback.

Harrison and Stanford (1989) describe the implementation of CML into a statistical methods course for first year students at the University of Queensland. No statistical analysis was performed on the results of the program, however their paper discussed some of the experiences and lessons learned. From their experiences they conclude that students liked CML because it allowed them to determine the pace at which they worked and because it provided feedback. Harrison and Stanford concluded by acknowledging that while it was difficult to provide precise studies of student performance from participation in the CML program, it was their judgment that student performance had been enhanced by the program.

Zakrzewski and Bull (1998) report staff reactions to the use of computer-based summative assessment as very favourable, although with some qualifications. Staff saw the main advantages as a reduction in marking effort, marking consistency and the fast availability and analysis of results. Concern was expressed about acquiring the skills to write items effectively to test analysis, synthesis and evaluation as well as comprehension, application and factual knowledge.

Bull (1999) reported some preliminary results of a new project investigating the implementation and evaluation of computer assisted assessment (CAA) throughout

the Higher Education section in the United Kingdom. CAA was the term used to encompass the use of "computers to deliver, mark and analyze assignments or examinations" (p. 12). Included in this definition was the use of optical mark readers for the collation and analysis of data gathered. Preliminary results from this national survey, conducted in January 1999, included identification of a PC-based closed network as the most common delivery mechanism, followed by almost equal numbers of academics using optical mark readers or the web to deliver assessment. The main advantages of CAA were seen to be the speed of feedback to students, statistical analysis of results and automatic marking of responses. Hardware and software reliability and availability as well as the difficulty of writing good items were seen as disadvantages. A large proportion of the assessment was summative.

## Implications for this Research

The consistent feature across the variety of examples that Black and Wiliam (1998a) include in their review article on classroom formative assessment is that "attention to formative assessment can lead to significant learning gains" (p. 17). While acknowledging that they have not found any negative reports, Black and Wiliam do not guarantee that these learning gains are irrespective of the "context and the particular approach adopted" (p. 17). The research in this thesis will address "context" by investigating formative assessment across a number of subject disciplines using the CML practice test as the "particular approach".

A second area where there is a scarcity of data is the use and evaluation of the effectiveness of computer-generated tests for formative assessment. In the literature to date there has not been a systematic assessment of the use of computer-based testing as a formative assessment tool, using quantitative methods across a range of subject disciplines. Available data are almost exclusively qualitative and subject specific. The research presented in this thesis uses largely quantititative methods over a range of subject disciplines.

The CML practice test is a computer generated multiple-choice test used for formative assessment. There are no reports in the literature on the use of the CML practice test with an approach identical to the one implemented at Curtin University of Technology, although work by Stanford and Cook (1987) has similarities. They used the same CML system to generate tests on which students were given three attempts to reach the predefined mastery level. Very few articles mention CML or computer-based testing for formative assessment at all. This research will investigate the benefits of the CML practice test across a number of different subject disciplines, both to increase the "context" in which this "particular approach" can be applied and also to counteract or confirm the opinion that the subject itself may be a determinant of the impact of the effectiveness of the assessment. While students will receive a mark on the practice test and be informed which items they answered incorrectly, the practice test will be a formative assessment tool and will not contribute to the student's final grade.

The two most important components of formative assessment are firstly, a perception by the student that a gap exists between their present and desired state of understanding and, secondly, the taking of some action by the student to close the gap. To be aware of the gap, the student must receive some form of feedback on performance. The nature of this feedback is central to formative assessment (Black & Wiliam, 1998a) and it is addressed next.

## Feedback

If learning is a process of mutual influence between learners and their environment then it must involve feedback (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991). While the important role of feedback in improving student performance has been recognised by educational researchers (Mory, 1992), there is considerable variation in the nature and function of feedback (Doig, 1999). It is not just getting feedback that is important, but also knowing how to use it (Laurillard, 1996).

The term feedback is relatively recent, being first used in the 1930s with process-control systems. It was used in the description of these systems primarily as an error detection mechanism which provided guidance to the system so that it could function effectively. Doig (1999) sees the process of feedback in these systems as having three key elements: it must move towards a known objective; it must be monitored in such a way that any diversion from this objective part of the system is detected; and it must have the ability to rectify any gap between what is happening and what should be happening. He uses the analogy of a heating system. The heating system keeps the room at a given temperature by monitoring the temperature. When the room becomes too hot or too cold the system rectifies the problem by adjusting the amount of heat generated. However, scientific systems have defined parameters but social systems do not and the notion of feedback becomes more involved.

Kulhavy (1977) used the term feedback in a generic sense to describe "any of the numerous procedures that are used to tell a learner if an instructional response is right or wrong" (p. 211). He saw feedback as a variable ranging along a continuum from a simple "Yes-No" format to the presentation of substantial corrective or remedial information which "identifies and corrects errors – or allows the learner to correct them" (p. 229). Other researchers use the term in a broader context. It is viewed as "informed dialogue for improvement" (Brown & Knight, 1994, p. 114), information given to students on their progress (Johnstone, Patterson, & Rubenstein, 1998), information provided to the performer of any action about performance (Black & Wiliam, 1998a) or as information on the gap between the student's present and desired level of performance (Sadler, 1989). Sadler worked from Ramaprasad's (1983) definition of feedback as being "information about the gap between the actual and reference level of a system parameter which is used to alter the gap in some way" (p. 4). Ramaprasad's definition mandates that the information about the gap between the actual and reference levels must be used to alter the gap. If it is merely recorded and not used then it is not feedback. Black and Wiliam (1998a) also address this notion of altering the gap when they refer to the role of feedback in formative assessment. They argue that feedback serves a formative function only

when used in diagnosis, i.e., "assessment is formative only when comparison of actual and reference levels yield information which is then used to alter the gap" (p. 53).

The extent to which feedback is effective depends on a number of factors, including whether it is externally or internally generated, what information it provides, how students receive it and what subsequent use they make of it. Butler and Winne (1995) refer to external and internal (self-generated) feedback while Sadler (1998) distinguishes between feedback students receive from an external source and student self-monitoring.

## External Feedback

Traditionally in education, the term feedback referred to external feedback. The two types of feedback considered were knowledge of results (KR) and knowledge of the correct response (KCR). Kulhavy (1977) reviewed research that examined the effects of feedback on learning from text. He sought to specify conditions in which feedback could be used with greatest instructional effectiveness. He concluded that feedback's most important instructional effect was to correct erroneous responses rather than to strengthen correct responses. He argued that the effectiveness of error correction depended on the student's original certainty of the response being correct, with the stronger effect being when the student was marked incorrect for a response they had been sure was correct (Kulhavy, 1977; Kulhavy & Stock, 1989).

Roper (1977) examined feedback in a computer-presented teaching program that required constructed responses in answer to items. The program consisted of an introduction that explained the program and how to use the computer followed by 36 items on basic statistical concepts. This was followed by a 25-item posttest. There were three groups each comprising twelve students. The first group received no feedback, the second was told whether a response was correct or incorrect and the third group received correct or incorrect plus the correct answer. Roper found that

feedback providing the correct answer was superior to feedback that only told the student the answer was correct or incorrect and this in turn was superior to the absence of feedback. He also concluded that feedback acted primarily to correct errors, and that the number of errors corrected was related to the amount of information provided by the feedback. This study by Roper, together with others reviewed by Kulhavy (1977), indicated that KR and KCR had negligible effect if given after a correct response, that is, feedback's most important effect was to correct erroneous responses rather than to strengthen responses that were correct.

Bangert-Drowns et al. (1991) concentrated on the effects of mediated intentional feedback in their meta-analysis that reviewed 58 effect sizes from 40 reports. Mediated feedback is feedback that is delivered in some fashion other than between teacher and student, for example, computer-based feedback. The review compared groups of students who received identical instruction, except that one group received feedback on answers to items and the other did not. The types of feedback included were right/wrong, supplying the correct answer to an incorrect response, requiring students to answer an item repeatedly until they got the correct answer and, lastly, an extended explanation of why an answer was correct or incorrect. Feedback type had a small relationship with effect size ($F[3,54]=2.58$, $p=0.06$), due mainly to low average effect size with right/wrong option only (ES=-0.08). When learners were either guided to or given the correct answer, the average effect size attributed to feedback was higher, 0.31 standard deviations. This pattern was reported both as a pattern of results for individual studies and also as an intrastudy finding. Results suggested that the effective feedback should not only indicate the correctness of the result but also provide correction where necessary. However, Bangert-Drowns et al. (1991) cited work by Schimmel where the amount of information supplied by the feedback was not significantly related to feedback effects. Schimmel's work was a meta-analysis on feedback in programmed and computer-based instruction.

The two major categories into which external feedback theories were originally placed were response-strengthening theories and error-correcting theories. The

response-strengthening theories came from the behaviourist tradition that asserted that the consequences associated with a response served to reinforce or suppress the same response in the future. Bangert-Drowns et al. (1991), Kulhavy (1977), Kulhavy and Stock (1989) and Roper (1977) all found that feedback's primary importance was the correction of errors, and hence their work supported the error-correcting theory.

While research generally confirms that when learners attend to external feedback they are more effective (Bangert-Drowns et al., 1991; Kulhavy & Stock, 1989), Crooks (1998) believes that the most effective form of feedback will depend on three key variables. He lists the correctness of the answer, the student's degree of confidence in the answer and the nature of the task as important. If the answer is correct the student needs to have that confirmed. If the incorrect answer is to a factual question, the most efficient form of feedback is to supply the correct answer (Phye, 1979). For questions involving higher level skills, more detailed feedback is desirable. However, the nature of the task may be different for different students because it has the potential to be affected by both student style and strategy use. In addition, the work of Turner reported by Butler and Winne (1994) found that students wanted different things from feedback, with some students requesting detailed explanations and others preferring much less detail. Black and Wiliam (1998a) refer to several studies that show that "innovations designed to strengthen the frequent feedback that students receive about their learning yields substantial learning gains" (p. 7).

Educational thinking about external feedback has moved from the position where it was identified with knowledge of results and was task directed, through a second phase of equating feedback with praise for effort, to the more recent research where the feedback is tailored to the specific assessment task and attention is given to the learner's response to a particular task (Sadler, 1998). Sadler argues that formative assessment includes both feedback and self-monitoring or internal feedback.

## Internal Feedback

Customarily feedback has been regarded as external feedback. However, Butler and Winne (1995) believe that feedback needs to be placed in the broader context of self-regulated learning (SRL), which they define as a "style of engaging with tasks" (p. 245) that uses goal setting, deliberation about strategies and the monitoring of effects. Boekaerts (1997) defines self–regulated learning as a "complex interactive process involving not only cognitive self-regulation but also motivational self-regulation" (p.161). In a similar vein, Carver and Scheier's (1990) model of self-regulated behaviour identified the importance of feedback that learners generate for themselves, that is, internal feedback, as they engage in tasks.

So, in addition to the identification of errors and assistance in correcting them, feedback has motivational effects and can guide learning strategies (Butler & Winne, 1995). It can offer information about the study process and the mastery of the learning goals (Martens & Dochy, 1997), can confirm student's conceptual understanding or beliefs, add information, overwrite incorrect elements, tune understanding and restructure schema (Butler & Winne, 1995). Students who generate their own internal feedback, either from external feedback provided to them after responding to test items or from information supplied from other sources, are better able to monitor their learning and change strategies to fit the learning situation.

Zimmerman (1990) summarises the three key features that are included in the description of students who self-regulate. Firstly, they are metacognitively, motivationally and behaviourally active participants in their own learning process. Secondly, they use a self-oriented feedback loop during learning (Carver & Scheier, 1981; Zimmerman, 1988, as cited by Zimmerman, 1990). This loop is described as a cyclic process in which students monitor the effectiveness of their learning methods and strategies. Responses vary from covert changes in self-perception to changes in behaviour such as the replacement of one learning strategy with another. Thirdly, the

descriptions of SRL all contain a motivational dimension, i.e., a description of how and why students choose to use a particular self-regulated process, strategy or response.

Self-regulated learners interpret academic tasks and, based on their interpretation, they set goals. Progress towards these goals is monitored and this, in turn, generates internal feedback. The internal feedback allows the student to reassess and modify their involvement with the task. It is this monitoring that is seen as essential because it generates internal feedback that can guide further action. "As a result of monitoring task engagement, students may alter knowledge and beliefs, which, in turn, might influence subsequent self-regulation" (Butler & Winne, 1995, p. 248). Butler and Winne use the term "might", because the student has to make the choice. The student has to reassess the situation when an impediment is encountered, and then either abandon the goal as too hard or decide that additional effort or modification of plans, or both, are required. At this point Carver and Scheier (1990) say that self-regulation has taken place.

Active engagement in the learning process yields increases in academic performance (Ames, 1984; Corno, 1986; Dweck, 1986; Paris & Oka, 1986; Zimmerman, 1989a). Therefore, self-regulated learners are typically high achievers (Ablard & Lipschultz, 1998; Zimmerman & Martinez-Pons, 1990). For example, in a study (Paterson, 1996) comparing students' achievement under conditions of self-regulated and traditional (teacher-regulated) study, two equivalent groups of high achieving year 12 biology students were matched using IQ and initial Biology achievement performance scores. Higher measures of reported self-regulation were significantly associated with higher academic performance scores.

Zimmerman (1994) suggests that a key condition for self-regulation is the availability of choice and control for the subjects. As well as choice of method of learning, students need choice regarding the use of their time and choice over their performance outcomes. He sees this final choice as involving more than a

willingness to participate, he sees it as involving "self-monitoring and self-modulating selected outcomes of one's performance" (p. 9). Research has shown that choice, and the resulting perceptions of control, are motivational variables that are significantly related to achievement (Clark, 1988).

Boekaerts (1997) suggests that self-regulated learners do not just develop but rather teachers must create appropriate learning environments. For those students who have not learnt to regulate their learning in relation to a particular domain external regulation is required. She draws a distinction between external regulation that leaves the learner little autonomy or responsibility for the learning process, and "scaffolding" that "captures the idea of a temporary support system that helps an individual during the initial period of gaining expertise" (p. 171). She reviewed several studies in which students were taught one or more cognitive strategies and provided with scaffolding. She concluded that when students were encouraged both to solve problems and to reflect on the solution process they could acquire metacognitive knowledge and skills. The outcome was higher performance on curricular tasks in the same content domain.

## Factors Influencing Feedback Reception

Factors that influence the reception of feedback can be catorgorised into those factors that are directly related to the feedback and those that are related to student characteristics. Direct factors include the timing of the feedback, the instruction type, the difficulty level of the work being assessed as well as early answer availability prior to the delivery of this intentional feedback. Student characteristics that may influence the reception of feedback include the learning style, strategy and student goals. Additionally, the nature of the task has the ability to advantage or disadvantage students (Crooks, 1988) depending on their learning style and the strategies they use in studying the material. Hence the role of the student is also important (Black & Wiliam, 1998a).

Direct Feedback Factors that May Influence Feedback Reception

Kulhavy (1977) argued that the introduction of a small delay between response and feedback increased the impact of the error-correcting feedback, in contradiction to the behavioural immediate reinforcement theory. Bangert-Drowns et al. (1991) agreed, in that they found that immediate feedback after a test gave lower average effects than delayed feedback. However, immediate feedback was used in almost all of the programmed instruction (where a series of frames are presented to the student who responds to a frame, receives feedback and then proceeds to the next frame) and computer-based instruction studies (23 out of 27 effect sizes), and hence was confounded by instruction type. Kulik and Kulik (1988) completed a meta-analysis on the findings of feedback timing. The results were mixed, however, immediate feedback was found to be superior to delayed feedback for classroom quizzes. Delayed feedback was found to be effective but only in somewhat contrived situations, such as list learning with stimulus word repetition. A key factor that appeared to influence these differences was whether or not the test items were identical to the earlier items that supplied the feedback. Crooks (1988), in his review of the impact of classroom evaluation practices on students, concluded that for most classroom situations immediate feedback is more beneficial than delayed feedback. However, he also concluded that as typical effect sizes are not large, the precise timing of feedback does not appear to be too critical unless it is "delayed so long that students have little motivation to pay close attention and learn from it" (p. 457).

Bangert-Drowns et al. (1991) found that, on average, direct feedback made a small positive contribution to achievement. However 18 of the 58 effect sizes calculated had a negative effect for feedback. As this is counter-intuitive the authors searched for mediating factors. Answer availability before a student had to commit to their own answer was strongly related to effect size ($F[1,56]=22.77$, $p=.0001$). In studies where looking ahead was impossible, feedback made an important contribution to achievement, 0.46 standard deviations. This answer availability, labeled "presearch availability" by Kulhavy (1977), describes "the ease with which a learner can locate

correct answers without first searching through or reading the lesson" (p. 217). Kulhavy concluded that presearch availability must be controlled for if feedback was to make a positive contribution to achievement.

Feedback fails to perform its facilitative role if the material is too difficult for learners because they will spend time guessing the correct answer and then try to associate the feedback with the item (Kulhavy, 1977). Bangert-Drowns et al. (1991) found a correlation of –.44 between task difficulty and the benefit of feedback suggesting that where error rates are high, learning from the feedback is difficult. However, they also concluded that the feedback would be more important when the context was more difficult and when the student is given fewer clues and supports.

Feedback effects could also be differentiated according to instruction type with programmed and computer-based instruction being less effective than text comprehension and conventional testing (Bangert-Drowns et al., 1991). The use of a pretest produced lower effect sizes suggesting that the pretest gave students the opportunity to review content, practice test item format and it generally acted as an advance organiser for the test to follow (Bangert-Drowns et al., 1991).

It was found that the two variables most strongly related to effect size (control for presearch availability and type of feedback, i.e. right/wrong, supplying the correct answer, repeatedly answering a question and an extended explanation) together account for about 50% of the variance in effect sizes (Bangert-Drowns et al. 1991).

Student Characteristics as Factors that May Influence Feedback Reception

Martens and Dochy (1997) cite research findings that show that the effects of feedback are not always consistent. As effects are sometimes positive, sometimes non-existent and sometimes negative, they postulated that the explanation may be the interaction between student characteristics and assessment, with students requiring different things from feedback. Factors that contribute to student use of

feedback include learning styles and strategies used in the study process as well as the goals students set for themselves.

Style and Strategy

There is considerable overlap in the literature between the terms learning style and cognitive style and between the terms learning style, learning strategy and learning approach. However, as statistically significant correlations with academic performance have been found for some approaches and styles (Sadler-Smith, 1997), they will be discussed briefly and placed in the context of CML multiple-choice item tests.

In an overview of style development, Rayner and Riding (1997) refer to the cognition-centered approach, which uses the term "cognitive style", and the learner-centered approach which uses "learning style". Rayner and Riding refer to three key headings under the cognition-centered approach: Wholist-Analytic, Verbal-Imagery, and a third heading integrating both of these dimensions. The Wholist-Analytic style dimension categorised individuals on their ability to process information as a whole or in parts while the Verbal-Imagery style dimension related to the individual's inclination to represent information verbally or in mental pictures (Riding & Cheema, 1991). For the learner-centered approach, Rayner and Riding (1997) divide key work into three style groups: process-based, preference-based and cognitive skills-based models.

Key terms common in the literature include "converging-diverging" and "surface-deep". Convergent thinkers are often cited as good performers on problems requiring one correct answer, while the divergent thinker is better suited to problems requiring several equally acceptable answers where the emphasis is on originality of response (Riding, 1997). As the item type used in CML tests is multiple-choice and requires one correct answer only, the convergent thinkers are more likely to perform well. The "surface-deep" label (Biggs, 1978; Marton & Saljo, 1976; Ramsden, 1985) is often used in relation to multiple-choice items. Marton and Saljo (1976) found

marked differences between individuals in the types of learning processes they engaged in. The learner with the deep-level processing focused on the "intentional content" of the material while the student with the surface-level processing attempted to memorise the text. Students expressed differing views about the author's possible intent in the text and a strong relationship was found between the levels of output and the levels of processing. Students adopting deep approaches perform well on evaluations doing at least as well on lower cognitive level questions as their surface-oriented peers and doing much better on higher level questions (Biggs, 1973; Svensson, 1977).

The concept of style is "always associated with individuality and is invariably used to describe an individual quality, form, activity or behaviour sustained over time" (Rayner & Riding 1997, p. 5). It is considered to be a fairly fixed characteristic of an individual (Riding & Cheema, 1991), to have a physiological basis (Riding, 1997), and is not critical when the task is simple but is more likely to be important when the learner is under pressure because the task is then perceived as difficult.

In contrast, a strategy or approach is considered to be a learned characteristic (Butler & Winne, 1995; Riding, 1997; Riding & Cheema, 1991) which may have an effect on a response to a particular task. The important difference that Biggs and Moore (1993) see between style and strategy is that styles are permanent personality characteristics that are displayed over a range of situations while strategies are modifiable and reflective of the student's current motivation and the teaching context.

Both style and strategy have the potential to influence students' reception of feedback and performance on a task. The effect of style for an individual will be positive or negative depending on the task (Riding, 1997) while the student's recall of content is highly related to the strategies used earlier in studying that content (Crooks, 1988).

## Influence of Style and Strategy on Student Use of Feedback

Both style and strategy may influence the perceived value of the feedback received and student recall of content, which has the potential to affect future marks (Brown & Knight, 1994). Using Entwistle's categories for approaches to learning, of 'deep', 'surface', 'strategic' and 'apathetic', Brown and Knight (1994) drew a link between the different learning styles and the different attitudes to feedback. The deep learning style links with the need for detailed comments, the surface style with general comments, the strategic with comments directly related to marks and hints to achieve better marks, and the apathetic style with those comments that encourage and boost confidence. In contrast, Martin (1999) believes that there are no deep learners or surface learners, just students who adopt different approaches in different circumstances.

## Student Goals as a Personal Characteristic Influencing Feedback Reception

Goals that students set for themselves influence their performance on particular tasks as well as the value they attach to feedback. In the research literature two contrasting achievement goal constructs have received most attention. They have been labeled differently by different authors: learning or performance goals (Dweck & Elliot, 1983), task-orientated or ego-orientated goals (Nicholls, 1984), and mastery or performance goals (Ames & Archer, 1987, 1988). In some investigations a third goal construct of work avoidance or academic alienation has been found (Nicholls, Cobb, Wood, Yackel & Patashnick, 1990; Nicholls, Patashnick & Nolan, 1985).

Those students who use learning, task-oriented or mastery goals, value learning for its own sake and seek to improve their level of understanding based on self-reference standards. They believe they can improve their ability by expending more effort (Dweck & Leggett, 1988) and focus on understanding material, even when performance is poor, therefore they adjust their effort after failure (Pintrich & De Groot, 1990). This effort adjustment implies self-regulation (Ablard & Lipschultz, 1998), therefore it is not surprising that these students tend to use SRL strategies

(Wolters, Yu & Pintrich, 1996, cited by Ablard & Lipschultz, 1998). Ames (1992) reviewed evidence indicating that classrooms that allow students autonomy and the ability to make their own decisions promote mastery orientations.

In contrast, those students selecting performance or ego-oriented goals seek to demonstrate high ability or performance in relation to the performance of others. They are more likely to view their abilities as stable straits that can be judged in relation to others (Dweck & Leggett, 1988). The third goal, academic alienation or work avoidance is characterised by a desire to complete the work with the minimum effort.

Meece (1994) refers to work by Schunk (1991), where the influence of goal setting on student achievement behaviour was examined. The research indicated that the goals students set for themselves could affect how students studied and what they remembered. Further studies (Graham & Golan, 1991; Nolen, 1988; Nolen & Haladyna, 1990; Pintrich & Garcia, 1991) indicate that learning-oriented students tend to use deep processing strategies while ego-oriented students are more likely to use surface level processing strategies such as memorising.

The goals students set for themselves can also influence their use of feedback. Research by Schutz, reviewed by Butler and Winne (1995), suggests that whether or not a student requested feedback on incorrect test items depended on the goals that the student had set for achieving certain percentages. If the student had reached his/her desired level of attainment, e.g. 50%, then there was no need to find out what the correct response was to the 50% of questions answered incorrectly. So, the effects of feedback were dependent on the student's overall goal rather than on feedback content.

Effective Feedback

To be effective, feedback must be related to a set of predetermined objectives and outcomes. However, as Doig (1999) notes, "educational theorists do not concur as to exactly how these features should be defined in the educational context" (p. 10). Butler and Winne (1995) argue that objectives and outcomes can only be defined by the individual learner. Slavin (1987), as an advocate of mastery learning, appears to have objectives and outcomes mandated through the curriculum and hence the same for all students while Sadler (1989) takes a middle position. He has objectives and outcomes defined by the curriculum but needed to be made explicit and transferable to students.

Rowntree (1987) sees effective feedback as enabling students to identify strengths and weaknesses so as to improve weak areas and build on strengths. For feedback to have any impact on learning the student must use the feedback to improve understanding (Johnstone et al., 1998). There is agreement among researchers that the effectiveness of feedback is dependent not only on its existence or absence but also on a multitude of other factors, including the type of feedback supplied, its timing and the context in which it is used. In all cases the learner has to take an active role and to improve learning the student needs to receive appropriate feedback, early and often (Angelo & Cross, 1993).

Crooks (1988) outlines three key ways in which the effectiveness of feedback can be enhanced. Firstly, it must focus students' attention on their progress in mastering the educational task. Secondly, it should be delivered while still relevant and, thirdly, it should be specific and related to need. On the third point, Crooks notes that knowledge of results is often sufficient and more detailed feedback is only required when necessary to help the student work through misconceptions of weaknesses in performance.

When feedback is directed towards self-esteem rather than towards the task it can have negative effects on attitudes and performance (Cameron & Pierce, 1994; Kulger & DeNisi, 1996). So, while feedback can promote learning it can also inhibit it if it encourages automatic or mindless responses (Salomon & Globerson, 1987).

## Implications of Feedback for this Research

Students receive immediate feedback at the completion of their CML test. This is a system parameter that is fixed with the mode of delivery used at Curtin University of Technology. The only flexibility available is the choice of feedback supplied to students. Students can either be told an answer is incorrect or alternatively told an answer is incorrect and given the correct answer. As this CML system provides only limited feedback it is important that students make the best use of it.

The literature in non CML areas supports the view that, when given these two options, the supplying of the correct answer is the more effective. This research will investigate the effect of both options. Consistent with current practice, students will be encouraged to use the feedback they receive to review their answers before they leave the CML Laboratory. It is hoped this will discourage any tendency to "mindlessness" and may also encourage students to use the external feedback supplied by the CML system as a trigger to promote the generation of internal feedback. This internal feedback could be important for the process of self-regulated learning.

In the literature, control and choice are key factors required for self-regulated learning. Students are provided with an opportunity to exercise some degree of control and choice by being given the opportunity to sit an optional practice test prior to their assessed CML test. While lecturers prescribe a time frame, usually a week or longer, students are free, firstly to choose to sit the test, and secondly to select a time that is suitable to them, hence, in a small way they create their own learning environment.

Learning style, strategy and chosen goals may influence students' performance as these factors have the potential to affect the use of the feedback supplied by the CML system. Additional factors that may have an impact on performance are sex and test anxiety. These will be addressed next.

## Sex of Students

Sex is a factor that may influence students' academic performance on tests. Performance on multiple-choice items, the use of SRL strategies and anxiety are possible areas where differences exist between males and females.

Multiple-choice items in traditional examinations disadvantage girls as a group (R.J.L. Murphy, 1982), however student age and the subject under consideration have also been given attention. Parker and Rennie (1995) cite a number of studies, in relation to English O-Level examinations, that have suggested that males have an advantage on multiple-choice items while females have the advantage on those questions requiring a longer response.

In UK national assessment, girls outperformed boys in English and Mathematics at ages 7, 11 and 14 (C. Woodhead, The Times, 6th March, 1996 cited by Stobart & Gipps, 1998). However at age 18+, in the A level examination, boys gain more higher grades than girls even in subjects in which girls did particularly well at 16 (Stobart & Gipps, 1998). In the UK Assessment of Performance Unit's (APU) science project (1980-1984), when students from the same curriculum backgrounds were compared, all gender gaps disappeared at age 15, except for those sections "Applying Physics Concepts" and "Making and Interpreting Observations" (P. Murphy, 1996). In addition, Murphy cites research in Thailand showing girls at age 16-18 years performing better than boys in chemistry and at least as well as them in physics. Also in relation to subject area, Parker and Rennie (1995) cite a study by Jovanovic et al. (1994) where it was found that the male advantage on multiple-

choice items was significant for a physical science topic but not for a biological science one.

Anderson (1989) examined the performance of men and women students on objective tests in Mathematics at university level and showed significant differences between the sexes. He did offer a partial explanation by the higher degree of abstentionism on questions by women (no penalty for a blank but minus one mark for an incorrect answer), however he found that women also performed worse than men in those areas where abstentionism was low.

In his meta-analysis designed to provide a statistical synthesis of research on test anxiety, Hembree (1988) found that females consistently showed higher levels of test anxiety than males. However, he did note that the peak years were grades 5-10 and then the sex differences declined through upper school and college. However, it is of note that the degree to which students are confident they will succeed influences their success in a stressful testing situation with boys tending to be more confident in these situations than girls (Adams, 1986, cited in Parker & Rennie, 1995).

While initially it was thought that computer anxiety would be a transitory phenomenon, research has indicated that it is becoming more prevalent (Brosnan, 1999). The phenomenon has been found to be universal (Weil & Rosen, 1995) with estimates of as many as 50% of students in higher education being computer-anxious (Rosen & Maquire, 1990). In addition, Brosnan (1999) cites research that finds females are more computer-anxious than males. However, findings across the literature seem to be contradictory. In some studies there is indication that females experience greater anxiety than males (Bozionelos, 1996; Koohang, 1987; Okebukola, 1993) while others indicate there is no difference (Cohen & Waugh, 1989; Honeyman & White, 1987; Pope-Davis & Twing, 1991).

Additional factors that may contribute to performance differences include SRL and the actual test-taking situation. In general, SRL is associated with academic achievement (Zimmerman & Martinez-Pons, 1990), and girls report more frequent use of SRL strategies than boys (Ablard & Lipschultz, 1998). Girls also have higher mastery goals than boys (Ablard & Lipschultz, 1998). The test-taking situation may also contribute to differential performance. Parker and Rennie (1995) reported evidence of males having an advantage in external, timed assessment situations, while females appear to have an advantage on school-based or classroom-based, cumulative assessment.

### Implications for this Research

As academic performance appears to be affected differently for males and females by item type, test type, SRL and anxiety, sex will be investigated in the context of CML testing at Curtin University of Technology. The CML system uses multiple-choice items so there is the possibility that male students will have an advantage. While researchers generally attribute performance differences on this item type to students in the younger age groups, work by Anderson (1989) questions this premise. Additionally, the actual testing situation while computer-based, has some similarities to the classroom-testing environment, so it may advantage females. The ability to monitor progress relative to goals, specifically SRL strategies, may advantage females, while anxiety, as a possible reason for a pattern of male advantage in a particular testing situation, may discriminate between the sexes. Anxiety is further discussed in the following section.

## Anxiety

The stress connected with performing significant tests is recognised as a factor influencing student performance. While this is loosely known as test anxiety, in reality it is a complex mix of different physiological and psychological factors that have the potential to affect the results students obtain on CML tests. This section

will review what test anxiety is thought to be, how it is assessed and how it may influence student performance.

## Background

The first hint that test situations may be stressful came from physiological assessments in a group of medical students. The students were found to have glycosuria (sugar in the urine) after the examination but not before (Folin, Demis & Smillie, 1914). The study of test anxiety has progressed from the psychoanalytic theory of the German investigators relating anxiety to traumatic childhood experiences, through the development of the first psychometric scale developed by Brown and his colleagues at the University of Chicago in the 1930s and 1940s (Spielberger & Vagg, 1995), to the present models. Theories are now categorised into three basic models: cognitive-attentional models, in which variables like worry and task-irrelevant thinking foster test anxiety and impair performance; dual deficit models, that include both cognitive skills and academic skills such as study habits; and a model based on social learning theory (Smith, Arnkoff & Wright, 1990).

Based on responses to a Test Anxiety Questionnaire (TAQ), Sarason and Mandler (1952) grouped students according to their level of test anxiety. Initially the low test-anxious students outperformed their high test-anxious colleagues but as testing continued these performance differences tended to disappear. The initial difference was explained on the basis of two psychological drives; task-directed and anxiety. The task-directed drive stimulated behaviour by the student to reduce the drive by completing the task while the anxiety drive stimulated both efforts to finish the task and task-irrelevant behaviour detrimental to performance. Albert and Haber (1960) subsequently labeled the drives facilitating and debilitating, respectively. In a further attempt to define test anxiety, Liebert and Morris (1967) performed a factor analysis on Sarason and Mandler's TAQ and proposed that debilitating test anxiety was itself bidimensional. They proposed a model with worry and emotionality as its two major components. "Worry is primarily cognitive concern about the consequences of

failing, the ability of others relative to one's own" (p. 975) while emotionality is a physiological reaction that includes feelings of tension and apprehension. Emotionality is similar to the state anxiety that Spielberger (1972) describes as an emotional state experienced during examinations, which varies as a function of the perceived threat to the individual.

Spielberger (1966, 1972) saw early studies as equating test anxiety with the physiological arousal associated with activation of the autonomic nervous system, while failing to account for individual differences in anxiety susceptibility that influence an individual's perception of a situation as threatening or not. He distinguished between the stress of the examination, the threat this poses to an individual, and the resulting state anxiety that may be evoked. He saw trait anxiety as the second type of anxiety that was a relatively stable indicator of individual difference in anxiety susceptibility as well as an estimator of the probability that anxiety will be experienced in stressful situations. Test anxiety was seen as a situation specific form of trait anxiety (Spielberger, Gonzales, Taylor, Algaze & Anton, 1978) with worry and emotionality as major components. Sarason (1984, 1988) expanded on the two dimensional model (Liebert & Morris, 1967; Spielberger, 1980) and divided the cognitive component into worry and test irrelevant thoughts, while emotionality was composed of tension and bodily symptoms.

Different operational models have been proposed for test anxiety. The interference model (Mandler & Sarason, 1952 ; Sarason, 1972; Wine, 1971) has test anxiety disturbing the recall of prior learning and thus degrading performance while the deficit model (Culler & Holahan, 1980; Tobias, 1985) has lower scores as a consequence of inadequate study habits or deficient test-taking skills on the part of the student.

In his meta-analysis, Hembree (1988) supported the interference model, noting that the behavioural and cognitive-behavioural treatments were effective in test anxiety

reduction while study skills training did not seem to be. However, a study investigating the relationship between test anxiety and test performance with $10^{th}$ graders in two Arab high schools suggested that the deficit and interference models complement each other (Birenbaum & Nasser, 1994).

## Assessment of Test Anxiety

Recent methods for assessing test anxiety use self-report scales focusing on frequency and intensity of reactions that students experience in the testing situation. Contemporary scales are seen as adhering to one of three major positions: trait, situational or interactionist. Trait theory suggests that an individual has a predisposition to experience anxiety which is stable and independent of specific situational characteristics of the environment. The situational approach focuses on characteristics of the test situation which elicit anxiety reactions while the interactional model combines both trait and situational factors.

In an attempt to provide some comparison between the models, Bedell and Marlowe (1995) examined three test anxiety scales, one representing each of the three major positions, as well as one trait anxiety scale. The scales used were Sarason's 1978 Test Anxiety Scale (TAS) as an example of trait orientation, the Suinn Test Anxiety Behavior Scale (STABS) (1969) as a situational example, and Endler et al.'s (1962) S-R Inventory of Anxiousness as the interactional orientation. The fourth measure selected for the study was the Trait Anxiety subscale of State-Trait Anxiety Inventory (STAI) (Spielberger, Gorsuch, & Lushene, 1970). The scales were designed for college students. The TAS uses true-false items dealing with physiological, emotional, cognitive, and behavioural reactions during test-taking situations. The STABS consists of 50 items describing behavioural situations related to test anxiety. The S-R Inventory of Anxiousness consists of 14, five-point rating scales used to evaluate response tendencies to anxiety provoking situations (e.g. entering a final examination). The Trait Anxiety consists of 20 items on a four-point

rating scale used to measure individual differences in the tendency to respond to anxiety in response to various situations.

Bedell and Marlowe (1995) investigated the interchangeability of the scales as well as their use as predictors of performance on tests. The degree of correlation between the test anxiety scales and the STAI Trait Anxiety subscales ranged from .45 to .59, indicating that they were not significantly different. A further correlation showed that S-R Inventory, STABS and STAI Trait Anxiety all performed equally well on test prediction performance. They did not, however, predict more than 10% of the variance. TAS was not a good predictor of test performance.

## Influence of Anxiety on Academic Performance

Scores on anxiety measures usually show negative correlations with performance on achievement tests (Anderson & Sauser, 1995; Culler & Holahan, 1980; Gross & Mastenbrook, 1980; Hembree, 1988; Hunsley, 1985; Mandler & Sarason, 1952; Paulman & Kennelly, 1984; Sarason, Davidson, Lighthall, Waite, & Ruebush, 1960; Williams, 1996; Zoller & Ben-Chaim, 1988), however, some researchers note that correlations do not give evidence of causality (Anastasi, 1976; Sarason, 1972).

Ball (1995) drew attention to two principles that illustrate the complexity of the relationship between anxiety and performance. The first involved task difficulty. He noted that a more difficult task was more likely to elicit undesirable anxiety effects while an easier task was more likely to have facilitative effects. The second principle was related to test taker attributes. When the test taker has expert skills or is highly intelligent, anxiety is more likely to enhance rather than inhibit performance. He also noted that the relationship between test anxiety and test performance may not be linear and that increasing test anxiety may enhance test performance up to a certain point, after which greater test anxiety lowers performance.

The relationship is a complex one. From his meta-analysis of 562 studies, Hembree (1988) concluded that better performance accompanies test anxiety reduction, however he also cited those researchers who found that treatment for test anxiety does not necessarily increase test performance. The issue is not clear-cut as types of treatment for test anxiety vary. There is also the suggestion that test anxiety may be exaggerated when students are uncomfortable with the test-taking situation (Zoller & Ben-Chaim, 1988). Findings with the Trait Anxiety Inventory developed by Spielberger (1980) demonstrated that worry was negatively correlated to test performance but emotionality was essentially uncorrelated. However, Hembree's research (1988) saw emotionality triggering worry. He concluded that test anxiety was composed of two primary factors: worry, which was the cognitive concern about one's performance, and emotionality that was the automatic response to the test situation. On investigation of cause and effect between these components, he showed that when treatment was cognitive it did not seem effective in test anxiety reduction, however behavioural treatments were considerably more effective. Also, these behavioural treatments reduced both emotionality and the worry component. Hence, he suggested that emotionality triggers worry and test anxiety is a behavioural construct. "Test anxiety causes poor performance. This conclusion follows from the finding that better performance accompanies TA reduction" (Hembree, 1988, p.74).

While it is thought that anxiety has a detrimental impact on performance generally, Brosnan (1999) cites work by Bloom and Hautaluoma (1990) and Modafsky and Kwon (1994) where anxiety has a specific detrimental impact on learning to use computers generally. "Individuals high in computer anxiety will ... under-perform in computer-based versions of assessment" (Brosnan, 1999, p. 49). He notes that this only becomes relevant when the assessment is computer-based and not when the assessment is paper-based.

As individuals' levels of computer anxiety impact upon assessments that involve computers (Brosnan, 1999), computer-based assessment may disadvantage specific

groups of students. Unfortunately results of different studies lack consistency. Worthington and Zhao (1999) attribute this to both the lack of a clearly defined definition of computer anxiety and the fact that computer technology has undergone considerable change. In the 1970s the perception of the computer was as a computing machine or giant calculator. The 1980s saw the introduction of the graphical user interface and computers becoming more user friendly. Recently the computer has changed into a communications device (Worthington & Zhao, 1999) that allows users to access information and interact with people without having to know programming languages. Hence, the reaction of an individual to computers is quite likely to have changed over this period.

Computer anxiety typically has been characterised as fear about working with computers. Computer anxious people are classified on the basis of physiological responses to computer use (Jay 1981; Kelley & Charness, 1995; Torkzadeh & Angulo, 1992) as well as cognitive responses such as negative thoughts about computers (Kelley & Charness, 1995; Rosen, Sears & Weil, 1993). It is possible that this computer anxiety could compound test anxiety in the CML testing situation.

### Summary and Implications for CML

As fear and anxiety reactions in an individual can be triggered by a plethora of experiences (Carlson & Wright, 1993), it is possible that the specific testing situation for this research has the potential to adversely affect some students. While scores on anxiety measures usually show negative correlations with performance on achievement tests, the degree to which a student is anxious depends on the student's predisposition and perception of a particular situation as a threat. Because the influence of anxiety on performance will depend on how the individual student perceives the testing situation, it is not possible to predict a priori the impact of test anxiety on student performance using the CML system. This research plans to look at the general concept of anxiety, both trait and state, and use the results of this to gauge student reaction to the CML testing situation in this higher education

environment. The preceding sections have drawn attention to student attributes that may influence performance and the next section will consider the CML practice test and the CML testing situation at Curtin University of Technology.

## CML Practice Test

Differential performance between individual students or groups of students may be due to factors connected with the actual test taking situation or the format of the test items. For assessment to be fair and equitable the assessment practice must be such that all students have the "opportunity to reveal what they know, understand and can do" (Parker & Rennie, 1995, p. 898).

As students would choose to sit an optional practice test only if they perceived it to be beneficial, it is essential that the information they receive from the practice test allows them to identify which portions of material require more study and which have been mastered. In a study of the role of the practice test in predictions about test performance on texts (Maki & Serra, 1992), the use of the practice test had negative results. Even with feedback, similar practice tests were not useful in increasing students' predictions of their test performance. The study found that when the practice test was not identical to the criterion test, feedback decreased the accuracy of the predictions. Advice was given against the use of the practice test as an aid to students in assessing which portions of a text they knew well and which they did not. However, the practice test was considered useful in helping students learn information.

Bangert-Drowns et al. (1991) suggested that the use of pretests act "as primitive advance organizers for the instruction to follow" (p. 233). They found that the use of pretests produced significantly lower effect sizes when groups receiving feedback and no feedback were compared on a posttest.

In a study by Dalton and Goodrum (1991), learners were assigned to one of three different groups, a full pretest, an adaptive pretest and no pretest. In the full pretest learners were given twenty items and were expected to complete all twenty items. Learners in the adaptive pretest group were given the same pretest as was given in the full pretest treatment, however, as soon as a learner missed a total of five items he or she was not allowed to finish the pretest. The results, on posttest data, showed that the mean of the adaptive pretest group was significantly greater than the means of both the full and no-pretest groups. The researchers concluded that the adaptive pretest which often was relatively short, as learners were exited when they missed the appropriate number of items, "seemed successful in heightening expectations and aiding the acquisition of lesson rules" (Dalton & Goodrum, 1991, p. 211). They believe the full pretest may have lost much of its benefit by decreasing motivation. The results of their study support the view that a well designed pretest is not necessarily an extensive test. The test should only be long enough to allow a mastery/nonmastery decision to be made.

## Summary and Implications for CML

The majority of the preceding literature has dealt with testing, but not with CML testing as there is very little research reported. Those articles that do refer to CML testing mainly use student opinion or test results in a single subject discipline. There are a few articles on formative assessment in the area but the context is different from that used at Curtin University of Technology. Existing work is mainly on summative assessment. The main focus of this research is formative assessment in the CML context.

As the type of assessment used at university level has the potential to influence student grades (Brown & Knight, 1994; Heywood, 1977; Nightingale et al., 1996; Webster, 1992), the contribution of formative testing to improve student performance needs to be investigated in the context of CML testing. The features of the CML system used for formative assessment are the possibility of a practice test,

the feedback to students and feedback to lecturers. Both test attributes and students' personal approach to learning have the potential to affect performance.

The CML system provides students with external feedback and staff encourage students to take time to examine errors before leaving the CML Laboratory. This external feedback may act to promote internal feedback and monitoring. This could be similar to the self-oriented feedback loop that Zimmerman (1989b) describes as a cyclic process in which students monitor the effectiveness of their learning methods and strategies. While the external feedback is potentially beneficial, both for the promotion of internal monitoring as well as the triggering of internal feedback, it may also provide the scaffolding Boekaerts (1997) refers to as a means of temporary support in the initial periods of gaining expertise. The feedback provided may also help students re-evaluate their grasp of the topics and encourage the setting of goals for future tests. If the provision of feedback can trigger the generation of internal feedback and self-monitoring then the literature points to improved test performance.

The reference to the pretest as an advance organiser (Bangert-Drowns et al., 1991) was in the context of the lowering of the effect of feedback as measured by posttest data, with both feedback and no-feedback groups benefiting from the pretest. However, this organisational function may well be a beneficial use of the CML practice test which preceded the assessed test.

From this literature survey it can been seen that two crucial factors that may influence academic performance on the CML test are student attributes and test attributes. The effects of two student attributes, sex and student anxiety are investigated, but students' cognitive style, learning strategy and goals are beyond the scope of the study. Several test attributes are considered. Firstly, if the practice test covers similar content to the assessed test, then feedback on performance may diagnose areas of weakness and assist students' revision for the assessed tests. It may contribute to an improvement in self-monitoring or alert students that they need

to modify their learning strategy. Secondly, the practice test allows students to familiarise themselves with the CML system, and may allay anxiety during the assessed tests. The research will examine how the practice test affects performance on later CML assessed tests for first year university students. It will do this by examining those aspects of the practice test that can be moderated, especially anything that may affect student reception of useful information.

Research findings favour a deep approach to learning and task-oriented goals. If any of the modifications that can be made to the practice test allow students to shift from a surface approach to a deep approach then gains will have been made. However, the research does not aim to measure these attributes. It will consider student marks on both formative and summative CML tests, and students' comments to questionnaires and interviews, as an indicator of change.

# CHAPTER 3

# METHOD

This chapter describes the research questions, research design and method, and the instruments used for the fourteen studies reported in this thesis. First the research questions are outlined and the research design used to answer them is explained. Explanations of the CML system and its use at Curtin are followed by an overview of the studies, which initially are grouped chronologically and then rearranged according to the research question they address. Details are given of surveys, questionnaires and interviews used for additional data collection. The chapter concludes with a general overview of the analysis and its limitations.

## Research Design

The general research question asks whether the practice test affects later CML performance and what characteristics make it most effective. Specifically, the study will investigate the following questions:

1. How does the practice test affect performance on later CML assessed tests for first year university students?

   1(a) What is the effect of the amount of content overlap between the practice test and the assessed test?

   1(b) What is the effect of the length (number of items asked) of the practice test and what length do students prefer?

   1(c) What is the effect of feedback on incorrect responses and how do students use the feedback?

2. What are the characteristics of the most effective CML practice test?

Factors investigated as possibly having a confounding effect on test performance are (i) subject discipline, (ii) familiarity with the CML system, (iii) ability, (iv) students' anxiety level, (v) sex, and (vi) students' response to CML. These factors were investigated in order to rule out explanations for changes in student performance that may otherwise have been attributed to benefits provided by the practice test.

The study was carried out in the context of CML at Curtin University of Technology. Because data were collected in a real testing environment, the research questions were addressed using a series of small studies, each focusing on one-semester units for first year students. Although the research questions could be answered more effectively using a hypothesis-testing, experimental design, it was not possible (and for ethical reasons not desirable) to manipulate the testing environment for students. Consequently, the research design takes an eclectic approach, using both quantitative and qualitative methods of data collection, attempting to accumulate sufficient data from several sources to provide answers to the research questions.

The approach to the analysis was to examine the effect of a CML practice test on performance and to attempt to isolate factors that contributed to this effect. Some practice tests considered covered the same content as the assessed test, used the same number of items and gave students the correct answer to any item they answered incorrectly. This is how the initial studies were set up. Others covered varying fractions of the content of the assessed test, varied in the number of items asked on the test and in some later studies students were told only whether an answer was correct or incorrect and were not given the correct answer.

The quantitative data were student marks on the computer generated practice and assessed tests with data taken directly from the CML system. Qualitative data were

collected at two levels: firstly, data were collected by student survey or self-report questionnaire and secondly via interview of individual students.

Surveys were used primarily to gauge student opinions about CML testing, the practice test, and the feedback that the system gave, as well as allowing students to comment on any aspect of the testing situation. Initial surveys collected demographic data as well as general comment about the CML testing process, but later surveys requested comment specific to the practice test, the use of the feedback given by the CML system and the desired length of the practice test. A self-report anxiety questionnaire was used with students who had a compulsory practice test, and again with a group from a different subject discipline who had an optional practice test. The results were used to investigate anxiety and its influence on performance, and whether anxiety affected student choice to sit the practice test. Finally, some students were interviewed to uncover any new issues which might contribute to the findings.

As the nature of this research was cumulative, some issues identified by students on early surveys were investigated in later surveys and interviews. Data for most groups came from both CML test marks and at least one survey, self-report questionnaire or interview, however, two groups were required only to respond to a survey on the use of the feedback facility, as additional information was considered necessary to test earlier findings. A few students from these two groups also participated in an interview.

## Context of the Research

Lecturers use the CML system in a variety of ways. Most use the CML tests as one component of the total unit assessment with marks typically contributing about 20% towards the final unit mark. The number of assessed tests generated by the CML system for a particular unit varies between one and five. Some lecturers allow an optional practice test before the first assessed test, but very few make this

compulsory and most lecturers do not implement a practice test at all. This research has been instrumental in promoting the use of the practice test and the use of the practice test is increasing.

## Items in the Testbanks

The majority of items used through the CML Laboratory have been devised by individual lecturers. The exceptions to this are those items that are used for Psychology 113, Psychology 114 and Education 102, which were obtained from the textbook publishers. The proportion of items used on an individual test varies quite considerably. Many lecturers have only sufficient items to select one from three, but some of the larger testbanks, including those previously mentioned as coming from the publishers and Human Biology 134, would typically select one item from ten or more items.

As the tests are only as good as the items in the testbank, the quality of items is always an important consideration. Prior to 1998, only the Economics 101 items had been coded to reflect the degree of difficulty of individual items. However, this research has been instrumental in promoting the benefits of testbank analysis and its accompanying examination of the suitability of items. Most of the units involved in this research have now had at least one testbank revision. However, some of these revisions happened after the initial groups of students sat tests. Both Psychology 113 and Psychology 114 had revisions of their respective testbanks after the first group of students reported in this research used the CML tests. Other groups, including Medical Imaging 131 and Accounting 100, had minor revisions to individual items. In all cases the revision of items was assisted by feedback provided by the testbank analysis report. An example report is provided in Appendix A.

Students' marks are based on the tests generated from these testbanks. Lecturers receive a report on their testbanks, based on students' responses, at the end of each semester but there is no guarantee that they act on this. The report gives the

frequency with which students correctly answered selected items. The question statistics are shown by question type, with multiple-choice items supplying information related to the number of times the item was answered correctly as well as the number of times various distracters were selected. The quality of items is a factor outside the control of this research and there is no independent group on whom the psychometric properties of the items could be checked. Hence, by using a range of different subject areas in this thesis, it is hoped to provide a more complete picture of CML testing even allowing that some testbanks contain a small number of items and some items will be of a better standard than others.

## Standard Procedures for CML Testing

Students come to the CML Laboratory at their booked time, are marked off the booking list and proceed to a terminal. They select the appropriate subdirectory for their particular subject, identify themselves to the system using both their identification number (ID) and password, and their test is randomly generated. A paper copy of the test is printed with the student ID, and students are required to produce photo identification to verify that they are collecting the correct test. The first page of a sample test is provided in Appendix B. Students complete the test under supervision in the CML Laboratory, then enter their answers on the terminal. Again students are required to identify themselves with both ID and password, and are then prompted for answers to their specific test. Before answers are finally marked, the student is able to review and change any answer. No feedback is given at this stage. This review and change process can be repeated as many times as students require. When students indicate they are satisfied with their answers the test is marked and immediate feedback given on incorrect responses. Students are encouraged to print this feedback, which consists of the correct response to any item the student answered incorrectly, and use this "error summary" in conjunction with their printed test paper to examine their errors. Appendix C contains a sample error summary. Students may take their error summary from the Laboratory, but test papers are retained so as not to compromise the test bank. While in the CML

Laboratory, writing is disallowed on the error summary, however students are able to query any item or answer and request that their test paper be sent to their lecturer for comment. In this case students usually make an appointment to meet with the lecturer to discuss the test.

## Overview of Studies Conducted

All groups whose CML testing program involved a practice test were considered for the study. Students enrolled in all but one unit were offered an optional CML practice test, the exception was a Psychology 113 unit where this test was compulsory. Surveys and self-report questionnaires administered as part of the research were non-compulsory and volunteers were called for the structured interviews. The shorter interviews, conducted in the CML Laboratory, when students received their printed error summary, were also voluntary.

Data from the CML system were used to answer all parts of Research Questions 1 and 2. Survey data contributed additional information about the effect of the length of the practice test and what length students prefer as well as the effect of feedback on incorrect responses and how students use the feedback. Survey data also addressed sex, familiarity with the CML system, student ability, subject discipline and students' response to CML. Self-report questionnaires addressed students' anxiety level as a factor having a possible effect on test performance as well as providing information that could be related to student choice to sit the practice test. Table 3.1 provides an overview of all studies reported in this thesis and is accompanied by a brief description of the studies in chronological order.

Table 3.1 Overview of Studies

| Study | Discipline | *n* | Time | Practice test condition | Instrument used | % AT content covered by PT | Ques. PT | Ques. AT |
|---|---|---|---|---|---|---|---|---|
| 1 | Economics 101 | 227 | Sem1 '97 | Optional with feedback | S | 100 | 20 | 20 |
| 2 | Economics 101 | 614 | Sem2 '97 | Optional with feedback | S | 100 | 20 | 20 |
| 3 | Psychology 113 | 396 | Sem1 '98 | Compulsory with feedback | Q | 50 | 15 | 30 |
| 4 | Instrumentation 213 | 42 | Sem1 '98 | Optional; limited feedback | - | 100 | 10 | 30 |
| 5 | Accounting 100 | 190 | Sem2 '98 | Optional with feedback | S, Q | 75 | 10 | 20 |
| 6 | Psychology 114 | 360 | Sem2 '98 | Optional with feedback | S | 100 | 15 | 30 |
| 7 | Psychology 113 | 39 | Sem2'98 | Optional with feedback | - | 25 | 20 | 80 |
| 8 | Psychology 113 | 557 | Sem1 '99 | Optional with feedback | - | 0 | 15 | NA |
| 9 (a) | Medical Imaging 131 | 53 | Sem1 '99 | a) First – optional with feedback | S, I | 100 | 10 | 27 |
| 9 (b) | | 51 | | b) Second – optional; limited feedback | | 100 | 10 | 30 |
| 10 (a) | Accounting 100 | 749 | Sem1 '99 | a) First – optional with feedback | S | 100 | 10 | 20 |
| 10 (b) | | 718 | | b) Second – optional with feedback | | 75 | 10 | 20 |
| 11 | Human Biology 134 | 155 | Sem1 '99 | - | S, I | - | - | - |
| 12 | Education 102 | 65 | Sem2 '99 | - | S, I | - | - | - |
| 13 | Psychology 114 | 360 | Sem 2 '99 | Optional with feedback | S | 100 | 15 | 30 |
| 14 | Medical Imaging 132 | 50 | Sem 2 '99 | Optional; limited feedback | I | 100 | 20 | 20 |

Note. PT = practice test, AT = assessed test, S= survey, Q = questionnaire, I = interview, NA = not applicable

Of the 14 studies, the CML components of 12 are included in the analysis of CML test results. In these studies students had a practice test prior to one or both of their CML assessed tests. Practice tests varied. They could draw the same number of items as the assessed test, across the same content range and according to the same test parameters. While not identical to the assessed test, due to the drawing of a number of items from an objective that contained more items than were required for selection, these practice tests mirrored the assessed test. Other practice tests drew fewer items than were selected on the assessed test but chose from the same content range. The third category of practice test drew items from some fraction only of the content covered by the assessed test.

In the remainder of this section each of the studies will be described briefly and in the subsequent section the studies are linked with the research questions. Some studies have multiple parts that may be included in later sections but are referred to in this section to give a total picture of the study. Students involved in these studies were all enrolled in first year one-semester units at Curtin University of Technology.

## Study 1

The purposes of this two-part study were to ascertain whether the CML practice test had any effect on student performance on the CML assessed test and to obtain student demographic information relating to sex and mode of study. Students in a first semester Economics 101 unit participated.

### CML Part

As part of their unit assessment students were required to take tests in the CML Laboratory with the option to sit a practice test prior to their first assessed test. Both the practice test and the assessed test selected 20 items from the same two modules. Hence, the practice test covered 100% of the content of the assessed test.

Student Information

Students were given the *Attitude to CML – Study 1* survey, which is contained in Appendix D, when they came to the CML Laboratory to sit their second CML assessed test. The purpose of the survey was to gauge student feeling towards computer managed learning assessment, both as a component of the overall unit mark and also as a measure of student satisfaction with laboratory procedures and test options that were able to be modified.

## Study 2

A different group of students was doing the same Economics 101 unit but they were studying it in second semester. The purpose of this two-part study was to investigate the influence of a practice test on subsequent CML assessed tests as well as to ascertain possible factors that influenced selection of the practice test. Student demographic information relating to sex and mode of study (full-time or part-time) was also sought.

CML Part

The CML component of Study 2 paralleled that of Study 1 with both the practice and assessed test selecting 20 items across two modules.

Student Information

The student survey, *Student Demographics – Study 2*, contained in Appendix E, was handed to students when they took their second CML assessed test in the CML Laboratory. Students were required to identify themselves as the information was to be linked with CML achievement. Questions requested students to identify their mode of study, sex and expectation of results as well as university entrance score.

## Study 3

This two-part study was with Psychology 113 students in first semester 1998. Part 1 of Study 3 was the CML component and Part 2 was a self-report questionnaire on anxiety. The purposes of the study were to ascertain whether students experienced anxiety that affected their performance on CML tests as well as to investigate the benefits of a practice test that covered only part of the content of the assessed test.

### CML Part

The compulsory practice test had 15 items drawn across two modules and the assessed test had 30 items, 15 drawn in a parallel fashion to the practice test, and 15 additional items covering new content. Thus, the practice test covered 50% of the content of the assessed test.

### Anxiety Questionnaire

The second part of this study investigated anxiety as a potential influence on performance on CML tests. Before the practice test, and again before the first assessed test, students were asked to complete the *Anxiety Questionnaire - Study 3*. The instrument used was the State-Trait Anxiety Inventory for Adults by Charles D. Spielberger, which is described later in this chapter and a copy is in Appendix F. Spielberger labelled this scale "Self-Evaluation Questionnaire", presumably so that he did not alert respondents to its purpose as a measure of anxiety, and this title appears on the version in Appendix F. The questionnaire is more accurately described as a self-report measure than a self-evaluation measure.

As it was anticipated that all students would sit the compulsory practice test, students were asked to complete the first self-report questionnaire when they came to the CML Laboratory for the practice test. They were asked to sign a form agreeing to participate in the research and then complete both the Trait-Anxiety and the State-Anxiety scales. The two scales were on opposite sides of one page. When

students came to the CML Laboratory to sit their first assessed test, they were again requested to complete the double-sided Anxiety Questionnaire.

## Study 4

Students taking Instrumentation 213 during semester 1, 1998, participated. This study had only a CML component. It investigated the influence of informing students whether their responses to items on the practice test were right or wrong but not supplying them with the correct response. This was described as the limited feedback condition.

### CML Part

Students were given the choice of a practice test before the first and only CML assessed test. The practice test consisted of 10 items and covered the same content range as the assessed test but selected one third as many items.

## Study 5

First year Accounting 100 students participated in this second semester unit. The purposes of this three-part study were, firstly, to investigate the potential influence of the practice test with respect to content coverage, length of the practice test and subject discipline, secondly, to ascertain whether anxiety was a factor affecting students' performance on tests or their choice to sit a practice test and, thirdly, to investigate students' attitude to the CML testing system and their use of the practice test.

### CML Part

While students had three CML assessed tests only the first was preceded by an optional practice test. The practice test was 10 items in length and covered modules 1 to 3. The assessed test was 20 items in length and drew items from modules 1, 2, 3

and 4. It drew three items from module 1, six each from modules 2 and 3 and five from module 4.

This practice test had half the number of items on the assessed test and covered 75% of the content range. Assessed tests 2 and 3 for this unit are not considered in this study.

Anxiety Questionnaire

This study used the same instrument for measurement of anxiety as that used in Study 3. Students sat the Trait-Anxiety scale in a Lecture and the State-Anxiety scale before their assessed test in the CML Laboratory using the *Anxiety Questionnaire – Study 5* which is in Appendix F. Anxiety was investigated for its influence on performance with students from a different subject discipline than those of Study 3.

Accounting students (Study 5) were asked to complete one Trait-Anxiety scale and one State-Anxiety scale. The Trait-Anxiety scale was completed in a lecture prior to the CML practice test. Students were addressed in a lecture with the Lecturer present and they were informed that participation was entirely optional. As the self-report form is usually a doubled-sided one containing the Trait-Anxiety scale on one side and the State-Anxiety scale on the other side, it was specifically printed with the Trait-Anxiety on one side of the paper only. Students completed this form in the Lecture Theatre. A second (repeat) lecture was held later in the day and the Lecturer handed the form to students with the same instructions that the first group had received. Students were asked to complete the State-Anxiety scale when they came to the CML Laboratory to sit their assessed test.

Student Information

Appendix G contains the survey, *CML Usage - Study 5*, which was used to determine students' attitude to CML as a potential influence on performance.

Students' attitudes were investigated as they indirectly have a bearing on the final research question regarding the most effective practice test as a test can not be effective if students choose not to sit it. This survey was handed to students in the CML Laboratory on completion of their final test. Questions were asked about the students' attitudes to the CML testing system as well as their use of the practice test. Those students who sat the practice test were asked how they prepared for the test and how they felt it helped their performance on the assessed test. Those who chose not to sit the practice test were asked why they did not.

## Study 6

This two-part study considered a second semester Psychology 114 unit that had a first semester Psychology 113 unit (Study 3) as its prerequisite, thus some students were also in Study 3. Part 1 looked at CML marks while Part 2 was a student survey. Prior exposure to CML, student ability and length of the practice test were investigated as well as CML test performance.

### CML Part

Students had an optional practice test before the first assessed test. Fifteen items were selected randomly across each of three modules so students received 5 items from each module. On the assessed test they received 10 items from each of the same three modules. So, the practice test covered the same content range as the assessed test but selected half as many items.

### Student Information

A student survey, *CML Usage – Study 6*, which investigated students' attitudes to CML, was Part 2 of this study. The survey is parallel to that used in Study 5 and is in Appendix H. Prior exposure to CML was considered because 85% of students had used the CML system in their first semester Psychology 113 unit (Study 3).

## Study 7

This second semester Psychology 113 study had only a CML component. The main purpose of this study was to investigate content coverage by the practice test as a factor affecting student performance on the assessed test.

### CML Part

The practice test had one quarter of the number of items and covered 25% of the content range of the assessed test. It was an optional practice test covering the same two modules as covered by the Study 3 practice test, however, instead of drawing 15 items from each module it drew 10 items from each. In Study 3 the assessed test covered four modules, but in Study 7 the assessed test covered eight modules and was 80 items in length.

## Study 8

Study 8 was the same Psychology 113 unit as Study 3 and Study 7 but in a different semester and with different students. This study addressed the issue of familiarity with the CML system as having an effect on student performance on the assessed test as the practice test covered different material from the assessed test.

### CML Part

The unit was run in a first semester, had an optional CML practice test and two CML assessed tests. The 15-item practice test covered different content from that examined by the 30-item assessed test, so there was 0% overlap of content. Results from the second assessed test are included in this study because by the time students sat the second assessed test, all had used the CML testing at least once previously.

## Study 9

Students in this three-part study were enrolled in a first semester Medical Imaging 131 unit that was part of a Medical Science course. The purpose of the CML component was to examine student performance on the assessed test according to whether the preceding practice test supplied the correct answer to an incorrect response or whether the answer was merely marked correct or incorrect (limited feedback). Student survey data were sought, after the second CML assessed test, to investigate students' use of both the practice test and the feedback supplied. Students also were asked their preference for mode of feedback. Finally, student interviews were used to seek additional information and to confirm information supplied by the survey.

### CML Part

Students had two CML assessed tests each preceded by an optional practice test. Study 9(a) reports the first practice and assessed tests while Study 9(b) investigated the second practice test and second assessed test.

Both practice tests contained 10 items drawn at random across the same range of content as the corresponding assessed test. The first practice test provided the correct answer for any incorrect response while the second practice test only allowed the student to know if they had answered the item incorrectly (limited feedback). Notes were held in the CML Laboratory and could be used for reference by the student when going through the error summary on completion of the test but before leaving the CML Laboratory.

### Student Information

Student survey data were obtained via the *Feedback Survey – Study 9* which is in Appendix I. Students were asked questions relating to their use of the practice test facility, the amount of study and revision they did for the various tests as well as their use of the feedback provided by the CML system. Students were able to

provide general comments regarding the information about incorrect answers as well as suggestions for improving the CML testing system. This survey was given to students on completion of their second assessed CML test.

Student Interviews

The final part of Study 9 was student interviews. Students were addressed in a lecture and asked if they could give some comments relating to the CML testing. The voluntary interviews investigated how students felt about the different feedback conditions as well as determining how they used the information supplied. Reasons for sitting or not sitting the practice tests were also considered as well as alternative test options, including an option that allowed students to attempt a test twice but use the higher mark to count towards the final grade. Only two students gave long interviews while a further three gave brief comments or a written response.

## Study 10

A group of first year undergraduate Accounting 100 students participated in this two-part study. Part 1 was the CML component while Part 2 was a student survey. Information was collected relating to length and content of the practice test as well as student choice of feedback mode. Study 10(a) focused on the first practice and assessed test and Study 10(b) focused on the second.

CML Part

Students were able to sit an optional CML practice test prior to either or both of their CML assessed tests. Both practice tests selected 10 items at random across the defined range of content while the assessed tests selected 20 items. The first practice test covered the same content covered by the first assessed test in Study 10(a), and the second practice test, Study 10(b), covered 80% of the content of its respective assessed test.

Student Information

The student survey, *Feedback Survey – Study 10*, found in Appendix J, paralleled that of Study 9. Study 10 included a larger group of students from a different subject discipline than those of Study 9. The survey varied on one question only, which related to a choice between the feedback and limited feedback situation. Full feedback was available on all tests in this study.

## Study 11

The purpose of this two-part study was to discover students' opinion about whether they thought it would be beneficial to withhold some feedback but to supply additional information in the CML Laboratory so that they could look up the correct answer to any items they answered incorrectly. They were also asked for information relating to their use of the error summary supplied and the amount of revision they did between subsequent attempts at a test. Students taking Human Biology 134 in second semester 1999 participated. Part 1 was a student survey handed to students after their first attempt at a CML assessed test while Part 2 comprised short student interviews conducted in the CML Laboratory.

Student Information

The CML component of the assessment for this unit allowed students to have two attempts at an assessed test but to take the higher mark. While the CML marks are not included in this analysis, students were asked questions, using the survey *Feedback Survey – Study 11*, in Appendix K, relating to their use of the feedback supplied and their opinion regarding the two feedback options.

Student Interviews

Student interviews were short and addressed similar questions to the survey. When students received their error summary from staff in the CML Laboratory they were asked their opinion of the CML tests and what they thought of the option of

supplying limited feedback only. They were also asked their opinion of the new arrangement that allowed them to have two attempts at each assessed test but count the higher mark only.

## Study 12

This two-part study comprised a survey and short unstructured student interviews. As Study 11 had collected feedback data that related to an assessed test rather than a practice test, Study 12 was set up to determine whether student opinions and use of feedback were similar when the questions related to the practice test. The second part of this study was semi-structured interviews. Second semester Education 102 students participated.

### Student Information

Students were given a survey, *Feedback Survey – Study 12*, in the CML Laboratory on completion of their first assessed test. A copy of this survey is in Appendix L.

### Student Interviews

Interviews were conducted in a similar manner to those with the Human Biology students of Study 11. This time students were responding to their use of feedback after the practice test. Some questions were prepared in advance while others arose in response to students' answers to previous questions.

## Study 13

Students in Study 13 were enrolled in a second semester Psychology 114 unit. This unit was also offered in the previous year and was the subject of Study 6. Study 13 was in two parts. Part 1 was the CML component and students received a practice test that covered the same content as the assessed test, but was half as long. The purpose of the survey in Part 2 was to further investigate feedback and students' use of it, and their preferred test length, as well as indirectly investigating item quality.

CML Part

The optional practice test covered the same content as the assessed test. The practice test was 15 items in length and the assessed test was 30 items. This testbank had been renewed since the previous year. Many items were eliminated from the testbank and the remaining items were coded into two difficulty categories. Half of each test was made up of the more difficult items.

Student Information

Appendix M contains the *Feedback Survey – Study 13*, which students were asked to complete after their assessed test. It was administered through the CML Laboratory. The survey was similar to the one used in Studies 11 and 12, but this one also asked students what length they preferred the practice test to be and why. It contained an identical question to one used in Study 6, which had evoked some negative comment regarding the quality of testbank items. The testbank had been renewed with problem items deleted and those remaining had been coded as easy or difficult. Test parameters were changed so that students received items from each category.

## Study 14

The final study was with a second semester group of Medical Imaging 132 students who were also part of Study 9 in the previous semester when they were studying Medical Imaging 131. The study was in two parts, a CML component and student interviews. The purpose of the CML component was to examine student performance when feedback was limited while the interviews sought information on students' use of the feedback supplied by the CML system.

CML Part

The optional practice test and the assessed test contained 30 items and they covered the same content. The practice test did not supply students with the correct answer for a item they answered incorrectly, however the Lecturer supplied notes that were

held in the CML Laboratory. Students were able to use these notes for reference after their test had been marked.

Student Interviews

In a similar fashion to Studies 11 and 12, students were asked for information related to their use of the feedback supplied by the CML system as well as their use of the practice test.

## How the Studies Address the Research Questions

This section will link each study with the research questions or with the additional factors investigated as potentially affecting either student performance or student choice to sit a practice test.

Research Question 1 asks how does the practice test affect performance on later CML assessed tests for first year university students? This question is addressed across all studies that include a CML component while the effect of the amount of content overlap between the practice test and the assessed test, as well as the effect of practice test length and feedback supplied, are addressed across different studies. Table 3.2 summarises this information as well as additional factors that have the potential to affect student performance.

Table 3.2 Overview of Studies by Research Questions and Confounding Factors Investigated

| | Research Question | | | | Confounding Factors | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Study | 1(a) Content | 1(b) Length | 1(c) Feedback | 2 | Anxiety | Sex | Response to CML | Familiarity with CML | Ability | Subject Discipline |
| 1 | ✔ | ✔ | | ✔ | | ✔ | ✔ | | | ✔ |
| 2 | ✔ | ✔ | | ✔ | | ✔ | | | ✔ | ✔ |
| 3 | ✔ | | | ✔ | ✔ | | | | | ✔ |
| 4 | ✔ | ✔ | | ✔ | | | | | | ✔ |
| 5 | ✔ | | | ✔ | ✔ | | ✔ | | ✔ | ✔ |
| 6 | ✔ | ✔ | | ✔ | | | ✔ | ✔ | ✔ | ✔ |
| 7 | ✔ | | | ✔ | | | | | | ✔ |
| 8 | ✔ | | | ✔ | | | | ✔ | | ✔ |
| 9 (a) | ✔ | ✔ | ✔ | ✔ | | ✔ | | | | ✔ |
| 9 (b) | ✔ | ✔ | ✔ | ✔ | | | | | | ✔ |
| 10 (a) | ✔ | ✔ | ✔ | ✔ | | ✔ | | | | ✔ |
| 10 (b) | ✔ | | ✔ | ✔ | | | | | | ✔ |
| 11 | | | ✔ | ✔ | | | ✔ | | | ✔ |
| 12 | | | ✔ | ✔ | | | ✔ | | | ✔ |
| 13 | ✔ | ✔ | ✔ | ✔ | | | ✔ | | | ✔ |
| 14 | ✔ | ✔ | ✔ | ✔ | | | | | | ✔ |

Research Question 1(a) concerning the effect of the content of the practice test is addressed by Studies 1-10 and 13 and 14 which have practice tests whose content covers a range from 0% to 100% of that on the assessed test. Where the practice test does not cover the entire content of the assessed test, the assessed test is divided into parts. Part A contains content previously examined by the practice test while Part B is new content. Student performance on the practice test is compared with student performance on the whole assessed test for those students who sat the practice test. Part A and Part B of the assessed test are then compared for both the practice test and non practice test groups.

Research Question 1(b) concerns the length of the practice test. This question is addressed by studies with practice tests where the content coverage is equivalent to that of the assessed test but where a smaller number of items is used. These include Studies 4, 6, 9(a), 9(b), 10(a) and 13. A comparison is made between these studies and Studies 1, 2, and 14 where the practice test reflects the assessed test both in content and length. Table 3.1 shows the number of items on the practice test and the number of items on the assessed test for these studies. Additional information was gained by student survey from Study 13 which included a question on preferred length of the practice test and reasons for the choice.

Research Question 1(c) concerns the effect of feedback given for incorrect responses and how students use the feedback. Feedback is considered in two ways with seven groups of first year undergraduate students. Three groups participated in the CML component and five groups completed surveys. One group was in both sections. For the three groups in the Medical Sciences area, Studies 4, 9(b) and 14, in which students were not supplied with the correct response to an incorrect answer on their CML optional practice test, students' marks on the CML practice and assessed tests are examined to determine whether students show improved performance. Survey results obtained from five groups of students, Studies 9, 10, 11, 12 and 13, are used for the purpose of investigating student use of feedback provided by the CML

system. Questions on the surveys asked students how they used the feedback differently when it does not include the correct answer, and what kind of feedback students prefer.

Research Question 2 asks what are the characteristics of the most effective CML practice test? This question is answered by a synthesis of all studies.

A number of factors are identified as having the potential to influence either students' choice to do a practice test or students' performance on CML tests. Studies 6, 8, 9(b), 10(b) and 14 explore familiarity with the CML system by comparing the performance of students who have previously used the CML system with those who had not. Student ability is examined in relation to Studies 2, 5 and 6. Three different ability measures are used. The first is students' university entry marks that were obtained using a survey, the second is the final non-CML mark for the unit that was obtained from the lecturer, while the third is the final mark in the prerequisite subject, also obtained from the lecturer. Student anxiety is investigated using the State-Trait Anxiety Inventory in Studies 3 and 5 while the potential effect of the sex of the student is addressed in Studies 1, 2, 9 and 10. Subject discipline is covered by all studies. To set CML testing in a broader context, students' response to CML was examined. This was subdivided into three areas: students' general attitude to CML, their use of the practice test and if and how students use the feedback supplied to incorrect responses. Survey questions answered by students in Studies 1, 5 and 6 addressed attitude to CML, while survey questions answered by students in Studies 5 and 6 addressed general use of the practice test. Student use of feedback was addressed by survey questions used in Studies 11, 12 and 13.

## Instrumentation

### Surveys

Several surveys were used to gauge students' attitudes to the CML system and specifically the practice test, as well as to obtain student demographic data primarily

to link these with choice to sit a practice test. Qualitative information regarding student use of feedback and preference for both the mode of feedback and the length of the practice test was sought to strengthen the data obtained from the CML system. A summary of the surveys used is in Table 3.3 and copies of surveys are provided in Appendices D to M.

Table 3.3 Surveys and Questionnaires.

| Purpose | Study | Appendix | Discipline | *n* | % return |
|---|---|---|---|---|---|
| Attitude to CML | 1 | D | Economics | 277 | 76.9 |
| | 5 | G | Accounting | 190 | 39.5 |
| | 6 | H | Psychology | 170 | 56.5 |
| Student demographics Sex; ability | 2 | E | Economics | 614 | 96.4 |
| | 9 | I | Medical Imaging | 53 | 83.0 |
| | 10 | J | Accounting | 749 | 85.7 |
| Use of the practice test | 5 | G | Accounting | 190 | 39.5 |
| | 6 | H | Psychology | 170 | 56.5 |
| Anxiety | 3 | F | Psychology | 323 | 100.0 |
| | | | | 376 | 85.6 |
| | 5 | F | Accounting | 190 | 53.7 |
| | | | | 190 | 91.5 |
| Feedback | 9 | I | Medical Imaging | 53 | 83.0 |
| | 10 | J | Accounting | 749 | 85.7 |
| | 11 | K | Human Biology | 155 | 87.7 |
| | 12 | L | Education | 65 | 76.9 |
| Feedback and length | 13 | M | Psychology | 360 | 94.0 |

Attitude to CML - Study 1

The *Attitude to CML - Study 1* survey (Appendix D) given to Economics 101 students had ten questions, with questions 1 to 3 asking for information concerning

mode of study, previous Economics study and sex. For Question 4, students were asked to select from given options to describe the best features of the CML system and to add other options if they wished. Other questions asked about the opportunity to resit a test, their preference for the allocation of 20% of the unit mark to CML and to suggest improvements for CML testing.

Student Demographics – Study 2

This seven-question survey, *Student Demographics - Study 2* (Appendix E), asked for information on mode of study, previous Economics study, sex, university entrance score and expected test mark as well as requiring a "Yes" or "No" response from students regarding whether they sat the practice test. Students were required to identify themselves, by student identification number, as the information was to be linked with CML achievement.

CML Usage – Study 5 and Study 6

Students' attitude to CML was investigated by the first three questions on this eight-question survey, *CML Usage - Study 5* (Appendix G), or *CML Usage – Study 6* (Appendix H), handed to students at the end of the unit. The first question asked students how much they liked CML as a form of assessment and allowed responses from 1, "hate it", to 5, "like it a lot". Questions 2 and 3 inquired how well students thought CML assessment measured their understanding of the unit and how difficult the CML system was to use. Again they were required to select a response from 1 to 5. Students were asked how much study they did for the CML practice test as well as how helpful they thought the test was. Both questions allowed students to respond on a 5-point scale. Students were asked how they prepared for the CML practice test and how they felt it helped their performance on the assessed test. Those who chose not to sit the practice test were asked why they did not. Students were able to add additional comments on the survey. They were not requested to give their name and were not followed up if they did not respond.

Feedback Survey – Study 9 and Study 10

Question response in the *Feedback Survey* used in Studies 9 and 10 (see Appendices I and J) was based on a 5-point rating scale, or required a "Yes" or "No" response. Questions asked whether the practice tests were useful as preparation for the assessed tests and whether being given the correct answer helped. Students selected from 1 "not at all" to 5 "a great deal". "Yes" or "No" answers were required to questions asking if students sat particular practice tests. Comments were requested if the amount of study after one practice test differed from that after the other. Study 9 students had an additional question relating to which feedback method they preferred.

Feedback Survey – Study 11, Study 12 and Study 13

The *Feedback Survey* for each of Studies 11, 12 and 13 (see Appendices L, M and N) required students to answer "Yes" or "No" to six questions about whether they used feedback from the CML system. Questions related to the usefulness of being given the correct answer, the use students made of this answer and their opinion regarding having to find the correct answer if it was not supplied. Students were also asked open-ended questions about their use of the feedback as preparation for future CML tests and their opinion about the feedback information supplied by the CML system when marking answers.

## Anxiety Questionnaires

The State-Trait Anxiety Inventory by Charles D. Spielberger (1983) was used in this study as it has been shown to be essentially equivalent, in predicting test performance, to other scales investigated. Secondly, it has the advantage that it is recommended by the Psychology Department at Curtin University of Technology, due to its reported reliability and validity. Both the Trait-Anxiety and State-Anxiety scales were used. The Trait-Anxiety scale measures a student's general disposition to anxiety (how the student feels generally) while the State-Anxiety scale measures

the level of anxiety to the particular situation (how the student feels at that particular moment, i.e., immediately prior to the test). Results from these scales were investigated in relation to student performance on both the CML practice and assessed tests to determine whether anxiety in this CML testing situation influences student performance adversely as suggested in the literature. Computer anxiety will not be specifically investigated but will be assumed to be included in students' general anxiety reaction.

Anxiety was examined in two studies (3 and 5) as a possible factor affecting choice to sit a practice test and performance on the assessed test. Students responded to a four-point scale with 1 corresponding to "not at all" and 4 "very much so" for the 20-statement State-Anxiety scale and 1 corresponding to "almost never" and 4 to "almost always" on the 20-statement Trait-Anxiety scale. Copies of the two scales are in Appendix F. As mentioned earlier, Spielberger labelled the Inventory as "Self-Evaluation Questionnaire". Students were required to identify themselves by their student identification number to enable their CML test scores to be linked to their Anxiety scores.

## Interviews

Interview were conducted with either an individual or small group with the purpose of testing information gained from the surveys as well as giving students an opportunity to add any new factors that were important to them but had not been previously requested. Table 3.4 summarises the number of students involved in interviews. The individual interviews were short and conducted through the CML Laboratory with students from Studies 11, 12 and 14. The longer, more in-depth interviews were conducted outside the CML Laboratory with two groups of Medical Imaging students from Study 9. Another student supplied written responses. The main questions addressed are listed below. Not all students responded specifically to each question.

Why did you sit a practice test?

Why didn't you sit a practice test?

Did you take the practice test seriously?

How did you feel about not having the correct answers to the practice test?

Did you revise everything between the practice and real test or only sections that the practice test showed you made errors in?

Do you know how to read the error summary?

Do you use the error summary for revision?

If you knew which modules referred to course sections, would it be useful?

Should the practice test and the real test be the same length? Should they test the same sections of the course? Why?

How would you feel about having two attempts at a test and then taking the higher mark?

Would you study more for the first test in this situation than if it were a practice?

Students participating in the short interviews were all responding to a request for information regarding their use of the CML system, however tests in the various units were set up in different ways.

The interviews were on an informal basis as students received either their error summary or returned their test paper prior to leaving the CML Laboratory. Students participating were asked a selection of questions including a request for a general comment on how they liked or disliked the testing process, how they prepared for the practice test in relation to the amount of study they did, how they used the information they received on the error summary as well as any additional information they cared to supply. As students were responding to a request for their opinion of the CML situation in a spontaneous manner, the questions followed their responses and most students only answered two or three questions. Students were interviewed on different days and different times of the day. However, as the short

interviews occurred in a one-week time frame, only those students who had a scheduled test for that particular week, whether this was lecturer or self-initiated, were included.

Table 3.4 Individual or Small Group Student Interviews

| Data collection method | Study | Discipline | *n* |
|---|---|---|---|
| Long interviews – individual or small group | 9 | Medical Imaging | 5 |
| Short individual interviews in the CML Laboratory | 11 | Human Biology | 13 |
| | 12 | Education | 5 |
| | 14 | Medical Imaging | 3 |

Short Interviews in Study 11

Students in the Human Biology 134 unit (Study 11) were interviewed after their first try at an assessed test about their use of the CML test, including benefits and problems associated with it. In addition, they were given the opportunity to make any other comment they wished regarding the CML testing in general. These students were allowed two tries at each of their six CML assessed tests with the higher mark then counting as part of the unit assessment. They could progress through the tests at their own pace and select in which order they sat a test. No test had a time limit. The CML system had been set up with a "challenge" option. This required the student to select which test they wished to sit so it was possible for a student to sit test 6 before any other test if they wanted to. It was also possible for a student to progress through the tests in order, 1 to 6. While students were allowed two tries at any test, these tries did not have to be sequential.

Short Interviews in Study 12

Students in this Education 102 unit (Study 12) were allowed two tries at each of their five CML assessed tests with the higher mark counting. They were also allowed multiple attempts at the practice test that preceded each assessed test. The

practice tests contained only half as many items as their respective assessed tests, though they did cover the same content range. These students also used the "challenge" option but tests had to be finished by the end of defined semester weeks. Students were responding to questions at an interview after attempting the assessed test. They responded to similar questions to those asked of students in Study 11 relating to their use of the CML test including benefits or problems encountered. They were also given the opportunity to comment on any aspect of the CML testing they wished to.

Short Interviews in Study 14

The Medical Imaging 132 students in Study 14 had two CML assessed tests, each preceded by an optional practice test. The error summary from the practice test informed students of incorrect responses but did not supply students with the correct answer, however, for the assessed tests students were supplied with the correct answer. Both practice tests were the same length and covered the same content range as the corresponding assessed test. Course notes were available in the CML Laboratory and staff actively encouraged this group of students to use them. Students participating in these interviews were responding to similar questions to those used with students from Studies 11 and 12 but were doing so after their first assessed test.

## Overview of Data Collection and Analysis

### Data Collection

For most studies, classes were visited by the researcher and the students advised of the general procedures used in the CML Laboratory. Students were given the opportunity to ask for any additional information or clarification they felt they needed. For the studies involving the Anxiety Questionnaire (Studies 3 and 5), students were informed that they would be asked to assist when they came to sit their CML test. The questionnaire was not described as an anxiety measure.

Lecturer approval was gained for the use of all surveys and students were given the survey through the CML Laboratory on completion of a test. Participation was optional, however, the response rate was usually good. Interviews were voluntary and were all conducted by the researcher. One class was approached through a lecture and students volunteered while other students were happy to be asked their opinions in the CML Laboratory after completion of a test.

The process of data collection for surveys and questionnaires was the same for all studies except Study 5. Students were asked to fill in surveys on completion of a test they sat through the CML Laboratory. The only difference to this was the Trait-Anxiety scale that Study 5 students were requested to complete in a lecture.

## Analysis

Differences in scores between groups were tested for statistical significance using *t* tests. The magnitude of the difference was examined using effect sizes where the effect size is calculated as the difference between means divided by the pooled standard deviation (Hedges, 1981). Effect sizes for correlated comparisons are calculated following Dunlap, Cortina, Vaslow and Burke (1996) and are interpreted in the same way. Emphasis was placed on effect sizes in the interpretation of findings rather than statistical significance because the sample sizes varied across the studies.

For the student surveys, responses to "Yes" or "No" questions, or those requiring students to select an option from a defined number of options, were tallied and expressed as a percentage. For those questions where an open response was required the responses were sorted into groups. The researcher devised the categories for these groups. Response by category is reported as a percentage of total responses received for the question.

Scoring for Spielberger's Anxiety Inventory followed the accompanying directions. Each item on the anxiety measure was given a weighted score of 1 to 4 with 4 indicating high anxiety. Specific items on each scale were reverse scored as they reflected anxiety-absent items. To calculate scores for the State-Anxiety and the Trait-Anxiety scales, the weighted scores for the 20 items of each scale were added. Respondents who missed one or two items on either scale were included. Their mean score was calculated for the items they answered and then this value was multiplied by 20 and rounded to the next highest whole number.

## Limitations

As this research took place in an operational CML testing laboratory there were constraints on data collection. The main limitation of the study was the need to use naturally occurring groups of students for whom permission to participate could be obtained. This meant it was not possible to control for extraneous variables such as sex, ability and anxiety, and it was necessary to replicate studies where possible. It was also necessary to collect data from different sources and by different methods to endeavour to build an understanding of the role of the practice test by excluding alternative explanations of the effects.

To investigate the influence of the content covered by the practice test on the assessed test, multiple studies were needed with the content on the practice test ranging from 0% to 100% of the content on the assessed test. It was not possible for a practice test to reflect only the latter part of an assessed test as the timing of the practice test in some units meant that the content would not have been covered. With the exception of Study 8, any practice test that did not reflect the total content covered by the assessed test reflected some percentage of the first part of the test.

While it would be optimal to generate practice tests that gave students immediate feedback on one half of the test and provided no feedback on the other half of the test, it is not possible to implement this with any of the commercially available CML

systems. As a compromise, one of the groups of students who had a practice test before each of two assessed tests was given feedback on the first practice test but not on the second.

## Overview of Results Chapters

The results obtained from this research are reported in Chapters 4 and 5. Chapter 4 specifically addresses the first research question, how the practice test affects performance on the CML assessed test, by considering the improvement of students' performance with the use of the practice test, the influence of the content and length of the practice test as well as the feedback supplied to the students. Subject discipline, familiarity with CML and ability are addressed as factors that could contribute to student choice to sit a practice test or performance on a test. All data contribute to determining the characteristics of the most effective CML practice test, which is the second research question. In Chapter 5 the student survey, questionnaire and interview data are used to address the three remaining factors identified as having potential to affect the benefits of the practice test. These are anxiety, sex and students' response to CML, including the use of the practice test.

# CHAPTER 4

# RESULTS OF CML TEST DATA

This chapter presents results of the practice and assessed tests using data obtained directly from the CML system. The data will enable a comparison of results from students who sat a practice test with results of those who did not. The results will specifically address the first research question, which asks how the practice test affects performance on later CML assessed tests for first year university students. They are used to consider the influence of the practice test on student performance on the assessed test, of the content and the length of the practice test as well as feedback provided. Additionally, the chapter will serve to create a general picture of the benefit or otherwise of the practice test across subject disciplines.

Students' results on the CML practice tests are reported in sections. Section 1 considers influence of the practice test on performance, Section 2 considers the influence of the content covered by the practice test on performance, Section 3 addresses practice test length and Section 4 refers to feedback provided by the CML system. Section 5 examines students' subject discipline, familiarity with the CML system and ability as factors that can contribute to choice to sit the practice test or performance on CML tests.

## Section 1 – Influence of the Practice Test on Performance

This section reports CML test results for Studies 1, 2, 4, 6, 9(a), 9(b), 10(a), 13 and 14 in which the practice test covered the same content as the assessed test. These same studies also contribute to Section 3 which addresses practice test length. The nature of the CML component of these studies is described in Table 3.1 (see p. 61).

Study 1

The first study was conducted with a population of 277 Economics 101 students and their results for the practice and assessed tests are reported in Table 4.1. The percentage means showed that those students who sat the optional practice test increased their mean mark from 56.65 to 71.78, which is a statistically significant difference (dependent $t$=12.66, $p$<.001) with a substantial effect size of 0.98. On the first assessed test, this group of students also performed better than those who had not sat the practice test. The difference was statistically significant (independent $t$=4.38, $p$<.001) with an effect size of 0.54.

Table 4.1 Mean Scores (%) on the CML Assessment for Economics 101 Students (Study 1)

| | Practice test group | | | Non practice test group | | | |
|---|---|---|---|---|---|---|---|
| Test | *n* | Mean | SD | *n* | Mean | SD | *t* |
| Practice 1 | 152 | 56.65 | 16.17 | | | | |
| Assessed 1 | 152 | 71.78 | 14.45 | 125 | 63.56 | 16.36 | 4.38* |

*p<.001

Study 2

Study 2 involved the same unit as Study 1, but was offered in second semester to different students. In this Economics 101 unit, 614 students were enrolled. CML test results are reported in Table 4.2. Again there was a statistically significant difference (dependent $t$=13.28, $p$<.001) between the practice test and assessed test mark for those who did the practice test. The mean mark increased from 62.18 to 72.72, and the effect size was 0.68. On the first assessed test, those students who had sat the practice test performed better than those who had not. The effect size was 0.33 and the difference was statistically significant (independent $t$=3.81, $p$<.001).

Table 4.2 Mean Scores (%) on the CML Assessment for Economics 101 Students (Study 2)

| | Practice test group | | | Non practice test group | | | |
|---|---|---|---|---|---|---|---|
| Test | *n* | Mean | SD | *n* | Mean | SD | *t* |
| Practice 1 | 417 | 62.18 | 15.34 | | | | |
| Assessed 1 | 417 | 72.72 | 15.66 | 197 | 67.56 | 15.68 | 3.81* |

*p<.001

Study 4

This study was conducted with a small group of 42 Instrumentation 213 students who were offered an optional practice test which covered the same content as the assessed test. On the practice test, students were informed if a particular answer was incorrect but they were not given the correct answer (limited feedback condition).

Table 4.3 reports the CML test results. Those students who sat the optional practice test increased their group mean mark from 46.21 to 75.48, which is a statistically significant difference (dependent $t$=9.78, $p$<.001) with a very large effect size of 2.05. They also performed better on the first assessed test than those who had not sat the practice test, but the effect size was small at 0.23 and the difference between groups was not statistically significant.

Table 4.3 Mean Scores (%) on the CML Assessment for Instrumentation 213 Students (Study 4)

| | Practice test group | | | Non practice test group | | | |
|---|---|---|---|---|---|---|---|
| Test | *n* | Mean | SD | *n* | Mean | SD | *t* |
| Practice 1 | 29 | 46.21 | 15.90 | | | | |
| Assessed 1 | 29 | 75.48 | 12.10 | 13 | 72.69 | 12.21 | 0.69 |

Study 6

Table 4.4 reports test results for a second semester population of 360 Psychology 114 students. Of these students 85% had completed Psychology 113, which was the basis of Study 3, so these students had previous CML experience. In Study 6 the

optional practice test covered the same content as the assessed test but selected half as many items. Again, those students who sat the practice test increased their mean mark. In this case the percentage means increased from 57.17 to 68.30, which is a statistically significant increase (dependent $t$=8.40, $p$<.001). The effect size of 0.79 is substantial. On the first assessed test, this group of students also performed better than those who had not sat the practice test. Although the difference was not great, it was statistically significant (independent $t$=2.24, $p$=.026) with a small effect size of 0.24.

Table 4.4 Mean Scores (%) on the CML Assessment for Psychology 114 Students (Study 6)

| | Practice test group | | | Non practice test group | | | |
|---|---|---|---|---|---|---|---|
| Test | *n* | Mean | SD | *n* | Mean | SD | *t* |
| Practice 1 | 147 | 57.17 | 15.11 | | | | |
| Assessed 1 | 147 | 68.30 | 13.38 | 213 | 64.99 | 13.98 | 2.24* |

*p=.026

Study 9

Of the 53 first year Medical Imaging 131 students who sat the first assessed test, 43 (81%) chose to sit the first optional practice test (Study 9(a)). The second assessed test was taken by 51 students of whom 19 (37%) chose to sit the second optional practice test (Study 9(b)). Table 4.5 reports results for the two practice tests and two assessed tests. The first practice test gave the student the correct answer for any incorrect response while the second practice test informed the student only that the answer was incorrect (limited feedback).

The group mean mark for those students who sat the first practice test increased from 58.60 to 73.79 on the first assessed test, while the group mean for those who sat the second practice test increased from 39.47 on the practice test to 60.74 on the assessed test. In both cases there was a statistically significant difference (dependent $t$ =7.53, $p$<.001; $t$=4.14, $p$=.001) with substantial effect sizes of 1.03 and 1.28, respectively. Also in both cases, the group of students who sat the practice test

performed better on the assessed test than the group who had not sat the practice test. The difference was statistically significant only for the second assessed test (independent $t$=2.60, $p$=.012) with an effect size of 0.75. The difference between groups on the first assessed test was not statistically significant (independent $t$=1.66, $p$=.104), although the effect size of 0.58 was moderate.

Table 4.5 Mean Scores (%) on the CML Assessment for Medical Imaging 131 Students (Study 9)

| Study | Test | Practice test group | | | Non practice test group | | | |
|---|---|---|---|---|---|---|---|---|
| | | *n* | Mean | SD | *n* | Mean | SD | *t* |
| 9(a) | Practice 1 | 43 | 58.60 | 16.41 | | | | |
| | Assessed 1 | 43 | 73.79 | 10.49 | 10 | 67.50 | 12.20 | 1.66* |
| 9(b) | Practice 2 | 19 | 39.47 | 19.00 | | | | |
| | Assessed 2 | 19 | 60.74 | 13.70 | 32 | 51.31 | 11.78 | 2.60** |

*p=.104, **p=.012

Study 10(a)

Table 4.6 reports test results for both the first practice and the first assessed test for Accounting 100 students (Study 10(a)). This first practice test covered the same content range as the first assessed test but used half as many items.

Of the 749 students who sat this first assessed test, 237 (32%) chose to sit the first optional practice test. The group mean mark for these students increased from 63.00 to 70.30, which is a statistically significant difference (dependent $t$=5.96, $p$<.001), with an effect size of 0.45. Those students who sat the practice test performed better on the assessed test than the group who had not. The difference was statistically significant (independent $t$=4.87, $p$<.001) with an effect size of 0.38. The results of the second practice and assessed tests are in Section 2 where the reported studies have practice tests that do not cover the same content as their corresponding assessed tests.

Table 4.6 Mean Scores (%) on the CML Assessment for Accounting 100 Students (Study 10(a))

| | Practice test group | | | Non practice test group | | | |
|---|---|---|---|---|---|---|---|
| Test | *n* | Mean | SD | *n* | Mean | SD | *t* |
| Practice 1 | 237 | 63.00 | 17.99 | | | | |
| Assessed 1 | 237 | 70.30 | 14.31 | 512 | 64.60 | 15.17 | 4.87* |

*$p<.001$

Study 13

Table 4.7 reports test results for a second semester population of 360 Psychology 114 students. The percentage means showed that those students who sat the practice test increased their mean mark from 64.61 to 77.57, which is a statistically significant increase (dependent $t=10.34$, $p<.001$) with a substantial effect size of 0.97. On the first assessed test, this group of students also performed better than those who had not sat the practice test. Although the difference was not large, with a small effect size of 0.28, it was statistically significant (independent $t=2.61$, $p=.009$).

Table 4.7 Mean Scores (%) on the CML Assessment for Psychology 114 Students (Study 13)

| | Practice test group | | | Non practice test group | | | |
|---|---|---|---|---|---|---|---|
| Test | *n* | Mean | SD | *n* | Mean | SD | *t* |
| Practice 1 | 178 | 64.61 | 15.02 | | | | |
| Assessed 1 | 178 | 77.57 | 11.47 | 182 | 74.25 | 12.63 | 2.61* |

*$p=.009$

Study 14

Medical Imaging 132 students in second semester participated in Study 14 and except for those students who had withdrawn from the course, these students are those of Study 9. Table 4.8 reports test results for the practice and the assessed test. Of the 50 students who sat the first assessed test, 22 (44%) chose to sit the first optional practice test. The group mean mark for these students increased from 46.82 to 71.82, a statistically significant difference (dependent $t=8.89$, $p<.001$) with a very

large effect size of 2.06. Again it was found that those students who sat the practice test performed better on the assessed test than the group who had not. The difference was statistically significant (independent $t$=2.38, $p$=.022) with an effect size of 0.68.

Table 4.8 Mean Scores (%) on the CML Assessment for Medical Imaging 132 Students (Study 14)

| Test | Practice test group | | | Non practice test group | | | |
|---|---|---|---|---|---|---|---|
| | *n* | Mean | SD | *n* | Mean | SD | *t* |
| Practice 1 | 22 | 46.82 | 12.59 | | | | |
| Assessed 1 | 22 | 71.82 | 11.60 | 28 | 63.57 | 12.61 | 2.38* |

*$p$=.022

Summary of Results for Section 1

Results of the studies reported in Section 1 are summarised in Table 4.9. All studies had practice tests that covered the same content as their respective assessed tests. Under the "practice test group" column, the reported effect sizes reflect the differences in mean scores between the practice test and the assessed test for the practice test groups. The effect sizes reported under the "Non practice test group" column refer to the difference between the practice test and non practice test groups on the assessed test.

Table 4.9 Effect sizes for studies covering the same content as the assessed test.

| Study | Subject | Practice test group | | | Non practice test group | |
|---|---|---|---|---|---|---|
| | | Mean PT | Mean AT | Effect Size | Mean AT | Effect Size |
| 1 | Economics 101 | 56.65 | 71.78 | 0.98 | 63.56 | 0.54 |
| 2 | Economics 101 | 62.18 | 72.72 | 0.68 | 67.56 | 0.33 |
| 4 | Instrumentation 213 | 46.21 | 75.48 | 2.05 | 72.69 | 0.23 |
| 6 | Psychology 114 | 57.17 | 68.30 | 0.79 | 64.99 | 0.24 |
| 9(a) | Medical Imaging 131 | 58.60 | 73.79 | 1.03 | 67.50 | 0.58 |
| 9(b) | Medical Imaging 131 | 39.47 | 60.74 | 1.28 | 51.31 | 0.75 |
| 10(a) | Accounting 100 | 63.00 | 70.30 | 0.45 | 64.60 | 0.38 |
| 13 | Psychology 114 | 64.61 | 77.57 | 0.97 | 74.25 | 0.28 |
| 14 | Medical Imaging 132 | 46.82 | 71.82 | 2.06 | 63.57 | 0.68 |

Note. PT = practice test; AT = assessed test

Table 4.9 shows that in all nine studies, those students who sat the practice test increased their performance from the practice test to the assessed test. Further, in all cases the practice test group performed better than the non practice test group on the assessed test. There was considerable variation in effect sizes, which averaged 1.14, for the difference between the practice and assessed test marks representing the improvement from the practice to the assessed test. The effect sizes between the practice test and non practice test groups were smaller, averaging 0.45, and in two cases, Studies 4 and 9(a), did not represent statistically significant differences. Overall, it appears that those students sitting the practice test performed at a higher level on the assessed test than those who did not.

## Section 2 - Influence of Content Covered in the Practice Test

This section brings together those studies where the practice test covered less content than was examined on the assessed test. The range of content covered in these practice tests varied from 0% to 80% of the content on the assessed test.

Study 3

Psychology 113, the prerequisite Psychology unit for Psychology 114 (Study 6), had a compulsory practice test but 53 of the 376 students failed to take it. As the practice test covered only half of the content of the assessed test, the results from the first assessed test were divided into two parts for analysis. Part A covers topics examined by the practice test, and Part B covers topics not previously examined. Table 4.10 reports the percentage mean scores on the different tests for the 323 Psychology students who took the practice test and the small group of 53 who did not. For the practice test group, the mean mark increased from 70.90 to 75.76 for the total assessed test, a statistically significant difference (dependent $t$=5.34, $p$<.001) with an effect size of 0.33. Further, these students performed better on Part A (topics which were included in the practice test) than on Part B, which was new work. The effect size for the difference between Part A and Part B was 0.24.

Those students who did not sit the practice test scored lower than those who did. This group also scored less well on Part B than Part A and although the magnitude of the difference was similar to the practice test group, it was not statistically significant (71.76 vs 68.00, $t$=1.35, $p$=0.18), a direct reflection of the small size of this group. The effect size was 0.21, similar to the effect size between Parts A and B for the practice test group. Therefore, in terms of amount of content covered, it seems unlikely that the advantage shown by the practice test group on Part A compared to Part B, is due to covering the same content on the practice test.

Table 4.10 Mean Scores (%) on the CML Assessment for Psychology 113 Students (Study 3)

| | Practice test group | | | Non practice test group | | |
|---|---|---|---|---|---|---|
| Test | *n* | Mean | SD | *n* | Mean | SD |
| Practice 1 (compulsory) | 323 | 70.90 | 16.44 | | | |
| Assessed 1 | 323 | 75.76 | 12.18 | 53 | 69.58 | 14.51 |
| Part A (same content) | 323 | 77.51 | 12.95 | 53 | 71.76 | 15.55 |
| Part B (new content) | 323 | 73.98 | 15.97 | 53 | 68.00 | 19.68 |

Study 5

Table 4.11 reports test results for 190 first year Accounting 100 students of whom 78 (41%) chose to sit the optional practice test. The group mean mark for those students who sat the optional practice test increased from 65.64 to 72.88, which is a statistically significant difference (dependent $t$=4.06, $p$<.001) with an effect size of 0.41. Further, this group of students performed better on the first assessed test than the group who had not sat the practice test. The difference was statistically significant (independent $t$=3.03, $p$=.003) with an effect size of 0.45.

In this study, the optional practice test covered three quarters of the content covered by the assessed test. As for Study 3, the results from the assessed test can be divided into two parts and the mean scores on these parts compared. Part A covers topics examined by the practice test, and Part B covers topics not previously examined.

Table 4.11 Mean Scores (%) on the CML Assessment for Accounting 100 Students (Study 5)

| | Practice test group | | | Non practice test group | | | |
|---|---|---|---|---|---|---|---|
| Test | *n* | Mean | SD | *n* | Mean | SD | *t* |
| Practice | 78 | 65.64 | 19.20 | | | | |
| Assessed | 78 | 72.88 | 15.76 | 112 | 65.58 | 16.76 | 3.03* |
| Part A (same content) | 78 | 75.90 | 17.95 | 112 | 68.81 | 16.96 | |
| Part B (new content) | 78 | 63.59 | 20.32 | 112 | 55.90 | 25.06 | |

*p<.01

The results in Table 4.11 show that the 78 students who sat the practice test performed better on Part A than on Part B. However, it can be seen that those students who did not sit the practice test also performed better on Part A than on Part B (68.81 vs 55.90). Both the practice test and the non practice test groups showed a statistically significant difference between their mean marks on Part A and Part B. The difference for the practice test group was statistically significant (dependent $t$ = 4.77, $p$<.001) with an effect size of 0.64 while for the non practice test group (dependent $t$ =6.18, $p$ <.001) a similar effect size of 0.58 was calculated. Thus, as for Study 3, there does not appear to be an advantage in covering the content in Part A in the practice test. Instead, in both studies it may be that students found Part B more difficult or perhaps students had less time to study the work as it was covered closer to the test time.

Study 7

A small group of 41 Psychology 113 students who were studying their first Psychology unit in semester 2 participated in this study. This unit was the same as that studied by different students in Study 3. The optional practice test covered 25% of the content of the assessed test. Again, test results are divided into Part A, the content covered by the practice test and Part B, the new work. Because nearly all of

this small group (39 out of 41 students) elected to sit the practice test only the practice test group will be considered.

There was no statistically significant difference between the mean scores on the practice test and the assessed test (dependent $t$=-0.74, $p$=.466), and the mean score was a little lower on the assessed test. However, there was a statistically significant difference between Part A and Part B of the assessed test (dependent $t$=4.18, $p$<.001) with an effect size of 0.68, with a higher score on Part A, the content covered in the practice test, as reported in Table 4.12. This result is consistent with those of Studies 3 and 5 in that students found the items on the content covered more recently to be more difficult.

Table 4.12 Mean Scores (%) on the CML Assessment for Psychology 113 Students (Study 7)

| | Practice test group | | |
|---|---|---|---|
| Test | *n* | Mean | SD |
| Practice 1 | 39 | 65.26 | 14.46 |
| Assessed 1 | 39 | 63.59 | 11.75 |
| Part A (same content) | 39 | 70.39 | 15.53 |
| Part B (new content) | 39 | 60.64 | 12.84 |

Study 8

Study 8 was conducted with a population of 557 Psychology 113 students who were offered an optional practice test which covered Psychology content but not content that would be used on the assessed test. This meant that the content overlap was zero.

This lecturer had used the CML system for previous Psychology practice tests (Study 3 and Study 7) and decided to allow students to do a practice test which, from her perspective, was merely for familiarity with the CML system. Those students who sat the practice test were made aware of the CML testing procedures and saw

the format that Psychology multiple-choice items were to take. The group mean mark for those students who sat the optional practice test increased from 72.08 to 75.77, which is a statistically significant difference (dependent $t$=4.40, $p$<.001) with an effect size of 0.26.

Table 4.13 Mean Scores (%) on the CML Assessment for Psychology 113 Students (Study 8)

| | Practice test group | | | Non practice test group | | | |
|---|---|---|---|---|---|---|---|
| Test | *n* | Mean | SD | *n* | Mean | SD | *t* |
| Practice 1 | 323 | 72.08 | 15.14 | | | | |
| Assessed 1 | 323 | 75.77 | 12.83 | 234 | 71.32 | 14.34 | 3.78* |
| Assessed 2 | 318 | 75.53 | 12.92 | 227 | 73.53 | 14.19 | 1.69 |

*$p$<.001

Although they were tested on content that was not examinable on the first assessed test, the students who took the practice test performed a little better than those who did not. Perhaps this prior exposure contributed to improved performance on the assessed test. By the time they sat their second assessed test, all students had CML exposure and there was very little difference in performance. Results are reported in Table 4.13. The percentage means show that those students who sat the optional practice test performed better on the first assessed test than those who had not sat the practice test. The difference was statistically significant (independent $t$=3.78, $p$<.001) with a small effect size of 0.33. The percentage means between the two groups on the second assessed test show no statistically significant difference (independent $t$ = 1.69, $p$=.092) and the effect size was 0.15.

Study 10(b)

Table 4.14 reports test results for the second practice and the second assessed test for a group of first year Accounting 100 students. The practice test covered 80% of the content of the assessed test but used half as many items. The assessed test was taken by 718 students of whom 203 (28%) chose to sit the optional practice test that preceded it. The group mean mark for those students who sat the practice test

increased from 66.31 to 69.61, which is a statistically significant difference (dependent $t$=2.54, $p$=.012) but has a small effect size of 0.17. Those students who sat the practice test performed better on the assessed test than the group who had not sat the practice test. The difference was statistically significant (independent $t$=4.85, $p$<.001) with an effect size of 0.40. Further, these students performed better on the 16 items on Part A (topics which were included in the practice test) than on the 4 items of Part B, which covered new work.

Both the practice test and the non practice test groups showed a statistically significant difference between Parts A and B of the test, suggesting that they found Part B harder, although the small number of items it contained (4 out of 20) means that this conclusion needs to be treated with some caution. The difference between Part A and Part B for the practice test group was statistically significant (dependent $t$ = 8.35, $p$<.001) with an effect size of 0.67 while for the non practice test group (dependent $t$ =11.56, $p$ <.001) the effect size was modest at 0.55.

Table 4.14 Mean Scores (%) on the CML Assessment for Accounting 100 Students (Study 10(b))

| | Practice test group | | | Non practice test group | | | |
|---|---|---|---|---|---|---|---|
| Test | *n* | Mean | SD | *n* | Mean | SD | *t* |
| Practice 2 | 203 | 66.31 | 20.67 | | | | |
| Assessed 2 | 203 | 69.61 | 16.60 | 515 | 62.47 | 18.19 | 4.85* |
| Part A (same content) | 203 | 72.97 | 16.77 | 515 | 65.34 | 18.76 | |
| Part B (new content) | 203 | 56.16 | 29.94 | 515 | 52.04 | 27.69 | |

*p<.001

## Summary of Results for Section 2

Except for students from Study 7, where the practice test covered the smallest percentage of the assessed test (25%), those students who sat the practice test

increased their score from the practice test to the assessed test, the same finding as for Section 1. In Study 8, where there was no content overlap, the practice test group did better than the non practice test group. Table 4.15 summarises practice and assessed test mean marks as well as marks on Part A and Part B of the assessed test. For those choosing to take the practice test, the mean effect size between the practice test and the assessed test, across the four complete tests, was 0.26. The mean effect sizes between the two parts of the assessed test, Part A (same content) and Part B (new content) were 0.56 for the practice test group and 0.45 for the non practice test group.

The number of items in Part B on each of these tests was different between studies which has made comparisons difficult. In Psychology (Study 3) the assessed test Part B contained 15 items, which was half the length of the test, and in Psychology (Study 7) it contained 60 items which was three quarters of the length of the test. Part B on the Accounting (Study 5) test contained only 5 of the 20 items, and the Accounting (Study 10) Part B test contained only 4 of the 20 items.

Table 4.15 Effect Sizes for Studies Covering Partial Content of Assessed Test

| Study | Subject | Group | Mean PT | Mean AT | Mean Part A | Mean Part B | Effect Size PT vs AT | Effect Size Pt A vs Pt B |
|---|---|---|---|---|---|---|---|---|
| 3 | Psychology | PT | 70.90 | 75.76 | 77.51 | 73.98 | 0.33 | 0.24 |
| | | Non PT | | | 71.76 | 68.00 | | 0.21 |
| 5 | Accounting | PT | 65.64 | 72.88 | 75.90 | 63.59 | 0.41 | 0.64 |
| | | Non PT | | 65.58 | 68.81 | 55.90 | | 0.58 |
| 7 | Psychology | PT | 65.26 | 63.59 | 70.39 | 60.64 | 0.13 | 0.68 |
| 10 (b) | Accounting | PT | 66.31 | 69.61 | 72.97 | 56.16 | 0.17 | 0.67 |
| | | Non PT | | 62.47 | 65.34 | 52.04 | | 0.55 |

Note. PT = practice test; AT= assessed test; Pt A = Part A (same content); Pt B = Part B (new content)

Study 8, involving the unit Psychology 113 was a separate case as the practice test covered 0% of the content of that assessed test. The practice test group outperformed the non practice test group on the first assessed test, however, after both groups had had exposure to the CML system and the type of items used there was no difference between the group mean marks on the second assessed test.

As in the previous section, it can be seen that those students who sat the practice test did better than those who did not. This conclusion is based on the comparison of results between the assessed test marks of those students who sat the practice test and those who did not. Part A and Part B differences do exist, however, they are in the same direction for both the practice test and non practice test groups. This suggests that the reason may be associated with some factor other than the practice test, perhaps difficulty level or, as it is more recent work, less time for learning it.

## Section 3 - Influence of the Length of the Practice Test

The effect of the length of the practice test was examined on those studies where the practice test and the assessed test covered the same content range, that is, amount of content covered is constant. These are Studies 1, 2, 4, 9(a), 9(b), 10(a), 13 and 14, and because the details of these studies were reported in Section 1, only the relevant results are summarised in this section. In addition, students' responses to the relevant question from the survey given in Study 13 are reported. Here students were asked about their preferred length of test.

### Summary of CML Results

Table 4.16 shows results from Studies 1, 2 and 14 where the practice test and the assessed test are of equal length, Studies 6, 10(a) and 13 where the practice test is half the length of the assessed test and three studies (9(a), 4, and 9(b)) where the practice test is less than half of the length of the assessed test. Under the "Effect size (PT/AT)" column the reported effect sizes reflect the differences in mean scores between the practice test and the assessed test for the practice test groups. The

following column, "Effect sizes by group", reports the differences in group mean scores on the assessed test for the practice and non practice test groups.

Table 4.16 Effect Sizes for Practice Tests of Varying Length.

| Study | Discipline | Length PT/AT | Effect size (PT/AT) | Effect size by group |
|---|---|---|---|---|
| 1 | Economics 101 | 1 | 0.98 | 0.54 |
| 2 | Economics 101 | 1 | 0.68 | 0.33 |
| 14 | Medical Imaging 132 | 1 | 2.06 | 0.68 |
| 6 | Psychology 114 | 1/2 | 0.78 | 0.24 |
| 10(a) | Accounting 100 | 1/2 | 0.45 | 0.38 |
| 13 | Psychology 114 | 1/2 | 0.97 | 0.28 |
| 9(a) | Medical Imaging 131 | 10/27 | 1.03 | 0.58 |
| 4 | Instrumentation 213 | 1/3 | 2.05 | 0.23 |
| 9(b) | Medical Imaging 131 | 1/3 | 1.28 | 0.75 |

Note. PT = practice test; AT = assessed test

In all cases, students who sat the practice test improved their performance on the assessed test, however it can be seen that effect sizes vary with no consistent pattern. Based on the reported data it appears that practice test length is not related to students' performance.

Student Opinion about the Length of the Practice Test

Psychology 114 students (Study 13) were asked by *Feedback Survey – Study 13* to give their preferred length for the practice test. They were able to select one of seven options ranging from 10 to 40 items (see Question 9 on the survey in Appendix M). The length of the assessed test was 30 items. Students were then requested to give a reason for their choice of practice test length. Of the 340 students who returned the survey, 333 (98%) responded to this question and the most preferred option, chosen by 202 (60.7%) students, was 30 items. The results are reported in Table 4.17.

Students were asked reasons for their choice of practice test length. Not all students replied and some put multiple reasons. Reasons were examined for three groups of students: those who selected less than 30 items, 30 items, or more than 30 items. For the 95 students who wanted the practice test to contain some number of items less than 30, 85 gave reasons. The most frequent reason (35/85) was that the practice test was only a sample. The next common response, 18/85, was a general or vague comment suggesting that the student's chosen number was a good number of items for a test but adding no additional reason. For the group who chose 30 items, 185 provided reasons. The most common reason (123/185) was that this meant the practice test replicated the assessed test. Two additional reasons were that a 30-item test provided a better sample of content (25/185) with items that had a similar weighting to those on the assessed test (24/185). Those students who wanted more than 30 items gave the provision of a better sample of content as their most frequent response (26/36).

Table 4.17 Preferred Length for a Practice Test (Study 13)

| Question | Response choices | *n* | *n* % |
|---|---|---|---|
| **Q9**. If the real test has 30 items and you are able to have some say in the design of the practice test, how many items would it have? | 10 | 3 | 0.9 |
| | 15 | 23 | 6.9 |
| | 20 | 57 | 17.1 |
| | 25 | 12 | 3.6 |
| | 30 | 202 | 60.7 |
| | 35 | 6 | 1.8 |
| | 40 | 30 | 9.0 |

Summary of Results for Section 3

Practice test length does not appear to be a factor related to student performance on the assessed test. In all cases investigated, except Study 7, the practice test group increased their mean mark for the assessed test. In Studies 1, 2 and 14, where the practice and assessed tests were the same length, the mean effect size was 1.24. For

Studies 6, 10(a) and 13, where the practice test was only half the length of the assessed test, the mean effect size was 0.73. For those practice tests that were shorter than half the length of their corresponding assessed test this effect size was 1.45. However, there was a considerable variation in each cluster. Student preference was for the practice test to contain the same number of items as the assessed test, with the main reason given that it was a better reflection of the assessed test.

## Section 4 - Influence of the Nature of Feedback

One Instrumentation 213 unit (Study 4) and two Medical Imaging units (Study 9(b) and Study 14) used a practice test where students were informed that an answer was incorrect but they were not given the correct response (limited feedback condition). Study 9 involved the first semester unit which is prerequisite to the semester 2 unit involved in Study 14. Fifty-three students participated in first semester with fifty continuing into second semester. The details for these three groups of students (Studies 4, 9(b) and 14) were reported earlier in this chapter. They were included with those studies that had a practice test that covered the same content as the assessed test otherwise feedback is not as relevant because it is based on different content.

### Summary of CML Test Results

The summary results for the practice tests in Studies 4, 9(b) and 14 where students received limited feedback and the practice test for Studies 1, 2, 6, 9(a), 10(a) and 13 where students received full feedback are presented in Table 4.18. The effect size is calculated on the difference between the group mean marks of the practice test and the corresponding assessed test. Clearly, those students who sat the optional practice test under the limited feedback condition increased their mean score on the assessed test. The differences were statistically significant in all cases (dependent $t$ =9.78, $p$<.001, $t$ = 4.14, $p$=.001, $t$ = 8.89, $p$<.001) with large effect sizes of 2.05, 1.28 and 2.06, respectively. Whereas Table 4.9 reports effect sizes for both practice and non practice test groups for studies that covered the same content as the assessed test,

Table 4.18 uses data from the practice test groups of these studies only and reports the data according to the feedback condition used. Reference to Table 4.18 shows that the largest effect sizes reported for practice and assessed tests covering the same content are from those groups that received limited feedback.

Table 4.18 Mean Scores (%) on the CML Assessment Component for Studies with Different Feedback Conditions.

| | Practice Test | | | Assessed Test | | | |
|---|---|---|---|---|---|---|---|
| Study | *n* | Mean | SD | *n* | Mean | SD | ES |
| | | | Limited Feedback Condition | | | | |
| 4 | 29 | 46.21 | 15.90 | 29 | 75.48 | 12.10 | 2.05 |
| 9(b) | 19 | 39.47 | 19.00 | 19 | 60.74 | 13.70 | 1.28 |
| 14 | 22 | 46.82 | 12.59 | 22 | 71.82 | 11.60 | 2.06 |
| | | | Full Feedback Condition | | | | |
| 1 | 152 | 56.65 | 16.17 | 152 | 71.78 | 14.45 | 0.98 |
| 2 | 417 | 62.18 | 15.34 | 417 | 72.72 | 15.66 | 0.68 |
| 6 | 147 | 57.17 | 15.11 | 147 | 68.30 | 13.38 | 0.79 |
| 9(a) | 43 | 58.60 | 16.41 | 43 | 73.79 | 10.49 | 1.03 |
| 10(a) | 237 | 63.00 | 17.99 | 237 | 70.30 | 14.31 | 0.45 |
| 13 | 178 | 64.61 | 15.02 | 178 | 77.57 | 11.47 | 0.97 |

Student Opinion about the Nature of Feedback

Medical Imaging 131 students (Study 9) and Accounting 100 students (Study 10) answered questions on their respective surveys *Feedback Survey - Study 9* (see Appendix I) and *Feedback Survey - Study 10* (see Appendix J) that related to their opinion on the usefulness of being given the correct answer to an incorrect response. They were also asked to comment on their use of this feedback information supplied by the CML system.

Surveys were returned by 44 (83%) Medical Imaging 131 students (Study 9) and 641 (86%) Accounting 100 students (Study 10). Students were asked whether being given the correct answer was helpful. Response choices were on a 5-point scale from 1, "not at all", to 5, "a great deal". Of the 35 Medical Imaging 131 students who replied, 11 (31%) said it helped a great deal and a further 13 students (37%) rated the response as a 3 or 4. A similar pattern was seen with the Accounting 100 students, with 46 (19%) of the students who responded selecting "a great deal" and a further 148 (62%) students selecting 3 or 4.

Of the 18 Medical Imaging 131 students who made a comment, 12 (66.7%) said that the correct answer alerted them to error areas while 34 (39.6%) of the Accounting 100 students said that the correct answer alerted them either to error areas or to important areas of content. A further 43 Accounting respondents (50%) referred specifically to a particular item and being alerted to their error for this item.

Only those Medical Imaging 131 students who sat both practice tests were asked which method was more useful – practice test 1 (P1) where the correct answer was supplied or practice test 2 (P2) where they were marked correct or incorrect, only. Of the 16 students who responded, 14 (87.5%) selected the feedback option that offered the correct answer to an incorrect response. When asked to comment on why one method of feedback was more useful, 16 comments were received with 10 being general comments stating a need for the correct answer.

Table 4.19 reports students' comments on their use of the information supplied on incorrect responses. In addition to alerting them to both general error areas and specific item errors, students commented on the need to be supplied with the correct answer and noted their desire to remove test papers from the CML Laboratory.

Table 4.19 Student Comments Across Studies 9 and 10

| *Please write any comments you have regarding the information about incorrect answers and how you use or don't use this information* | | |
|---|---|---|
| Comment | Medical Imaging (Study 9) $n$=27 | Accounting (Study 10) $n$=327 |
| General positive: very helpful; positive for correct answer | 12 (44.4%) | 77 (23.5%) |
| Want test paper back | 9 (33.3%) | 91 (27.8%) |
| Alerts to error areas | 6 (22.2%) | 42 (12.8%) |
| Alerts to error in specific item | | 66 (20.2%) |

Summary of Results for Section 4

The results reported show that student performance increases from the practice test to the assessed test even when feedback is limited, in fact the effect sizes are the largest in the studies where this was the case. However, the two Medical Imaging groups were essentially the same students in their first and second semester and the Instrumentation unit is from the Medical Science area, that is, a similar subject discipline, which may be a confounding factor. Students are in favor of receiving full feedback rather than just being told whether they are right or wrong. Even so, the CML results do not show better performance when the correct answer is given.

## Section 5 - Factors Influencing Choice to Sit a Practice Test

Use of the practice test by subject discipline, familiarity with the CML system and ability were examined using data taken directly from the CML system. These are three of the six factors identified as possibly having an effect on test performance. (The others, students' anxiety level, sex, and response to CML, are considered in the next chapter.) They were investigated as possible alternative explanations for changes in student performance between the practice test and the assessed test. Students' subject discipline and prior use of CML testing are discussed together as factors affecting student choice to sit a practice test. Studies are listed separately

with reference to previous use of the CML system. Study 3 is excluded because the practice test was compulsory.

Use of the Practice Test by Subject Discipline

A summary of the studies which included an optional practice test and the percentage of students electing to sit it is provided in Table 4.20. The results show that, when given the choice to take an optional practice test, an average of 54% of students choose to do so, however this ranges between 28% and 95%. If the small group of students in Study 7 is excluded from this calculation, because they were strongly advised to sit the practice test, the range is from 28% to 81%.

Table 4.20 Selection of Optional Practice Test by Subject Discipline.

| | | | Practice Test | |
|---|---|---|---|---|
| Study | Subject | Previous use of CML | Yes | No |
| 1 | Economics 101 | No | 152 (55%) | 125 (45%) |
| 2 | Economics 101 | No | 417 (68%) | 197 (32%) |
| 4 | Instrumentation 213 | No | 29 (69%) | 13 (31%) |
| 5[a] | Accounting 100 | Some | 78 (41%) | 112 (59%) |
| 6[b] | Psychology 114 | Yes | 109 (37%) | 184 (63%) |
| | | No | 38 (54%) | 32 (46%) |
| 7 | Psychology 113 | No | 39 (95%) | 2 (5%) |
| 8 | Psychology 113 | No | 323 (58%) | 234 (42%) |
| 9(a) | Medical Imaging 131 | No | 43 (81%) | 10 (19%) |
| 9(b)[c] | Medical Imaging 131 | Yes | 19 (37%) | 32 (63%) |
| 10(a) | Accounting 100 | No | 237 (32%) | 512 (68%) |
| 10(b)[c] | Accounting 100 | Yes | 203 (28%) | 515 (72%) |
| 13[a] | Psychology 114 | Some | 178 (49%) | 182 (51%) |
| 14[c] | Medical Imaging 132 | Yes | 22 (44%) | 28 (56%) |

Note. [a] some participation - unknown percentage, [b] 85% of this group had prior CML exposure, [c] 100% of group had prior CML exposure

It was thought possible that subject discipline may be contributing to students' choice to sit a practice test but there is no clear trend. Both Accounting groups (Studies 5 and 10) have a small percentage of students sitting the test, however survey data revealed that at least some of the students in Study 5 had used the system previously. The percentage of students with prior use appears to be very small as this unit did not have a prerequisite unit that used the CML system and the unit itself is a first semester core unit. Those students who used the system previously may have been repeating students. Lecturer encouragement may also be contributing as shown by the results from the small Psychology group (Study 7) where the lecturer strongly encouraged participation. Those groups who had used the system previously also had low participation rates. In Studies 9 and 10 where students had two practice tests, fewer students sat the second one.

Familiarity with the CML System

Table 4.20 shows that participation rates of less than 50% were obtained from students in Studies 5, 6, 9(b), 10(a), 10(b), 13 and 14. Of these groups, only Accounting 100 students in Study 10(a) had not used the CML system previously. In the second semester Psychology 114 group (Study 6), 85% of students had also been enrolled in the prerequisite Psychology 113 (Study 3) in the preceding semester. They had prior exposure to CML because they had two assessed tests based on CML in Psychology 113. Of these students, 109 (37%) sat the optional practice test in Psychology 114 and 184 (63%) did not. Of the 70 remaining students who had not done the prerequisite Psychology 113 unit in the preceding semester, 38 (54%) sat the practice test and 32 (46%) did not. That is, a higher proportion of students sat the practice test if they had not experienced the CML system before. Medical Imaging students (Study 9(b)) and Accounting students (Study 10(b)) had prior exposure to CML tests as this was the second practice test in each unit and students had previously sat at least the first assessed test. Some students may also have sat the first practice test. Study 14 students were doing a second semester Medical Imaging 132 unit and all students had completed Medical Imaging 131 (Study 9) in the previous semester.

Across Studies 6, 9(b), 10(b) and 14, where students' previous participation rates are known, only 31% of students who had used the CML system for a previous test chose to sit the practice test. It is possible that prior exposure to CML may actually deter these students from sitting a practice test in a subsequent unit or before a second assessed test in the same unit.

Student Ability

Three performance measures which reflected student ability were investigated as possible factors influencing students' choice to sit the practice test. It could be that more able students chose to sit the practice test and this could explain their superior performance on the assessed test. Data on University Entrance Score were obtained from a survey administered through the CML Laboratory for the second semester Economics 101 group (Study 2). For both the Accounting 100 group (Study 5) and the second semester Psychology 114 group (Study 6) relevant marks were obtained from lecturers. These were the final examination mark for the unit and final examination mark for the prerequisite unit, respectively.

Students enrolled in a second semester first year Economics 100 unit (Study 2) were handed a survey, *Student Demographics – Study 2* (see Appendix E), when they took a CML test in the CML laboratory. Students were required to identify themselves by student identification number so that the information could be linked with CML achievement. Three hundred and eighty three (64.7%) of the 592 students who handed in the questionnaire answered the question related to their university entrance score. They were not required to give an exact score but to select which of three bands contained their score. It is likely that not all of the students in this group gained admittance to the university on the basis of an entrance score because there was a proportion of overseas students in the group. Based on analysis of this sample of 383 students, no association was found between university entrance score and students' choice to do the optional practice test (chi-square 2.80, p=.246).

In Study 5 with first year Accounting 100 students, the final examination for the semester was the performance measure used as the indicator of student ability. This examination was worth 50% of the total unit mark, covered the entire course and was not run on the CML system. Using this performance measure as a criterion, there was no difference in the mean mark on the final examination between those students who elected to sit the practice test and those who did not (63.9% vs 60.0%, $t$=1.49, $p$=.14).

Most of the students enrolled in Psychology 114 (Study 6) had completed the prerequisite, Psychology 113 (Study 3), in the previous semester. For these students only, the performance measure investigated was the final non CML examination mark for Psychology 113, which was worth 40% of the unit mark. There was no statistically significant difference between those students who sat the optional practice (n=109) in the second semester unit and those who did not (n=184), based on the final non CML examination mark for their previous Psychology unit (69.5% vs 66.8%, $t$=1.69, $p$=.09).

## Summary of Results for Section 5

The comparisons in this section suggest firstly, that there is no strong evidence that students from some subject disciplines are more likely to take the practice test than students from another. Secondly, if students have had previous experience with CML, they are less likely to select a practice test. Thirdly, on the basis of the three different ability measures used in three separate studies, no evidence was found to suggest that student ability affects their choice to sit a practice test.

## Overall Summary

All of the studies, except Study 7, demonstrate that those students who complete a practice test score higher marks on the subsequent assessed test. This effect is seen for students from a range of units in Economics, Psychology, Medical Sciences and Accounting.

The practice tests in the two Economics 100 units (Study 1 and Study 2), Instrumentation 213 (Study 4), Psychology 114 (Study 6), Medical Imaging 131 (Study 9), Accounting 100 (Study 10(a)), Psychology114 (Study 13) and Medical Imaging 132 (Study 14) covered 100% of the content of the assessed test. In all cases, the practice test group improved their performance from the practice test to the assessed test. This change was statistically significant in all cases, with an average effect size of 1.14. This group also outperformed the non practice test group on the assessed test. Again this difference was statistically significant, except for Study 4, with an average effect size of 0.45.

For Accounting 100 (Study 10(b) – practice test 2), Accounting 100 (Study 5), Psychology 113 (Study 3), Psychology 113 (Study 7), and Psychology 113 (Study 8) the practice test covered respectively 80%, 75%, 50%, 25%, and 0% of the content of the assessed test. Except for those students of Study 7, students who sat the practice test improved their performance from the practice test to the assessed test. Where there was a non practice test group for comparison, the practice test group performed better on the assessed test than the non practice test group. Students from both practice test and non practice test groups performed better on Part A of the assessed test, that part reflected by the practice test, than on Part B which was new content. On the evidence available, it does not seem that the amount of content overlap has a measurable effect on performance on the assessed test. It may be that the new content covered was more difficult or students had less time to study it.

Length of the practice test did not appear to be a factor affecting performance on the assessed test, however, students expressed a preference that the practice and assessed tests be of identical length because this was considered to be more reflective of the "real" test. The choice to sit a practice test did not appear to be influenced by students' subject discipline, although prior use of CML testing did lead to a smaller percentage of students electing to sit the practice test. Student ability also did not appear to affect students' choice to sit the practice test.

# CHAPTER 5

# RESULTS OF SURVEY AND INTERVIEW DATA

The general research question asks whether the practice test affects later CML performance and what are its most effective characteristics. This chapter will use data obtained by student survey, questionnaire or student interview to address the three factors identified as having potential to affect the benefits of the CML practice test. Students' anxiety level, sex and response to CML are examined as possible explanations for changes in student performance that may otherwise be attributed to benefits provided by the practice test. In addition, students' use of the practice test, particularly their use of the feedback it provides, is reported to present a clearer picture of the benefits of the practice test.

The chapter is divided into 3 sections. Section 1 reports data from questionnaires relating to anxiety, Section 2 addresses sex as it influences students' choice to sit a practice test, mark on the assessed test and reported anxiety level, while Section 3 reports students' response to CML and their use of the practice test. This third section uses both student surveys and interviews to report students' attitude to CML generally, in order to give a broad picture. It also reports students' use of the practice test and of the feedback supplied by the CML system. Finally, the main interview questions and answers are summarised, including illustrative quotes.

## Section 1 - Anxiety

Students' levels of anxiety were investigated in relation to CML test performance. The instrument used was the State-Trait Anxiety Inventory for Adults by Charles D. Spielberger (1983) (see Appendix F). It comprises separate self-report scales for measuring state and trait anxiety. The State-Anxiety scale has 20 statements that are used to assess how the person feels "right now" and the Trait-Anxiety scale has 20

statements that assess how the person feels "generally". These scales were used in Studies 3 and 5.

Study 3

Students in this Psychology 113 unit were asked to complete the Spielberger State-Trait Anxiety Inventory before the compulsory practice test and again before the first assessed test. Note that the inventory was labeled by Spielberger as "Self-Evaluation Questionnaire" (see Appendix F). Table 5.1 reports the results for the entire group. It can be seen that students were a little more anxious on the State-Anxiety scale before the assessed test, but there were no differences on the other scores.

Table 5.1 Mean Scores (%) and Standard Deviation for Anxiety Measures (Study 3)

| Task | *n* | Mean | SD |
|---|---|---|---|
| Trait (pre practice test) | 323 | 41.19 | 9.09 |
| State (pre practice test) | 328 | 41.81 | 11.02 |
| Trait (pre assessed test) | 320 | 40.68 | 9.49 |
| State (pre assessed test) | 322 | 45.43 | 11.99 |

In order to examine whether anxiety was related to later performance on the CML assessed test, Pearson's product moment correlation was calculated. When the sample was considered as an entire group, performance on the first assessed test had a negligible correlation with both the level of the students' state anxiety (r =-.046, p=.417) and trait anxiety (r=-.035, p=.538). Because not all students sat the practice test, even though it was compulsory, a further analysis treated the group in two parts – those students who sat the practice test and those students who did not. For the group of students who sat the practice test, the correlation between the students' trait anxiety before the practice test and performance on it was negligible (r=-.066, p=.242), and so were the correlations between performance on the first assessed test and students' state anxiety (r=-.027, p=.655) and trait anxiety (r=.008, p=.899). For the group of students who did not sit the practice test, there was no relationship

between state anxiety and performance on the assessed test (r =-.128, p=.402) or between trait anxiety and performance on the assessed test (r=-.188, p=.215).

Figures 1 to 7 represent the anxiety data on scatter plots in order to rule out the possibility that the negligible Pearson product-moment correlation might be due to a curvilinear relationship. Clearly this is not the case. All graphs indicate no relationship.

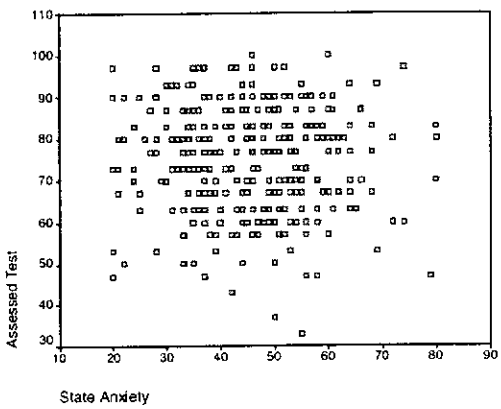


Figure 1.Assessed test mark and state anxiety for Psychology students (Study 3)

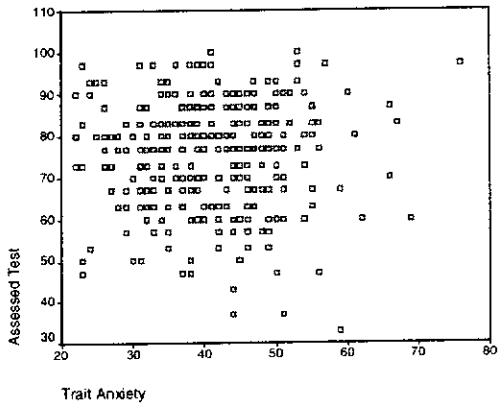


Figure 2. Assessed test mark and trait anxiety for Psychology students (Study 3)

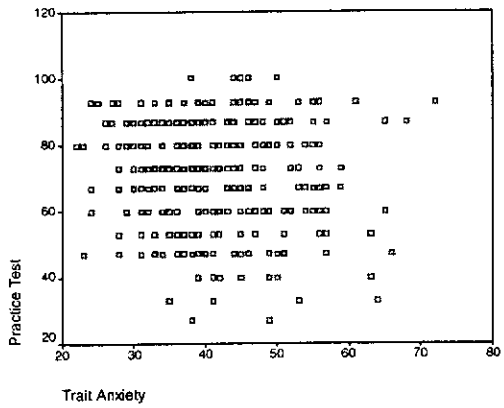


Figure 3. Practice test mark and trait anxiety for practice test group (Study 3)

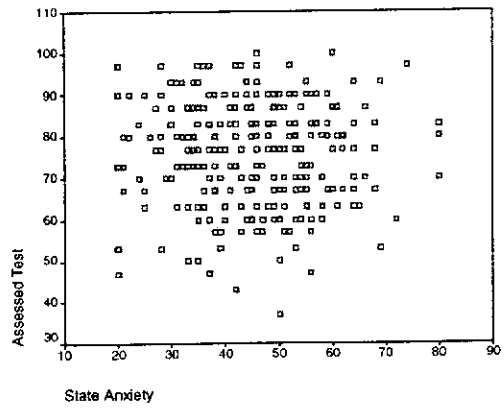


Figure 4. Assessed test mark and state anxiety for practice test group (Study 3)

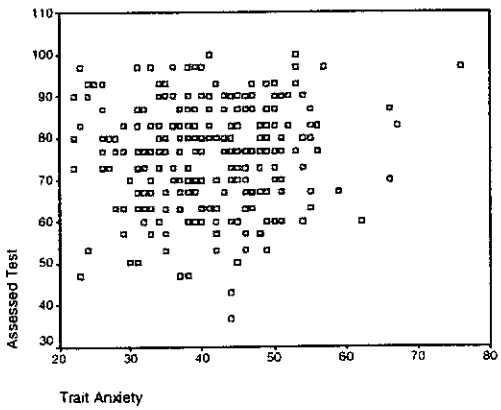


Figure 5. Assessed test mark and trait anxiety for practice test group (Study 3)

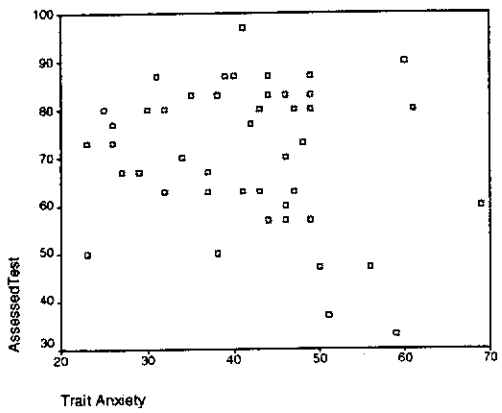


Figure 6. Assessed test mark and trait anxiety for non practice test group (Study 3)

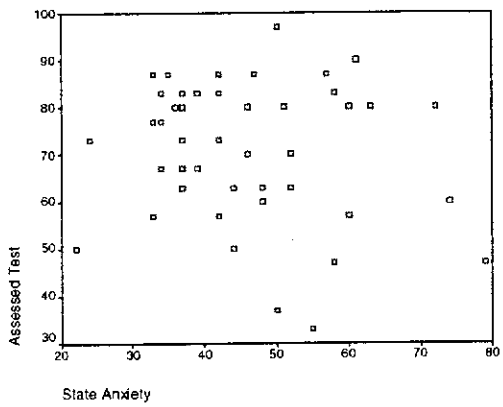


Figure 7. Assessed test mark and state anxiety for non practice test group (Study 3)

## Study 5

This group of Accounting 100 students also used Spielberger's State-Trait Anxiety Inventory in a parallel situation to Study 3. However, during a lecture, students were asked to complete the scale that measured trait anxiety prior to the practice test being available. Prior to the first assessed test, students were asked to complete the State-Anxiety scale.

Table 5.2 reports the results of the 102 students who completed the Trait-Anxiety scale and the 169 students who completed the State-Anxiety scale. No difference was found between those students who sat the practice test and those who did not, either on the measure of their state or trait anxiety.

Table 5.2 Mean Scores for State and Trait Anxiety Measures (Study 5)

| | Practice test group | | | Non practice test group | | | |
|---|---|---|---|---|---|---|---|
| Task | *n* | Mean | SD | *n* | Mean | SD | *t* |
| Trait-Anxiety (pre practice test) | 58 | 43.02 | 8.76 | 44 | 42.34 | 7.65 | .41 |
| State-Anxiety (pre assessed test) | 70 | 43.70 | 11.38 | 99 | 43.45 | 9.41 | .15 |

Pearson's product moment coefficient was again calculated to determine whether anxiety was related to later performance on the CML assessed test. In parallel to Study 3 this group was also treated in two parts – those students who sat the optional practice test and those students who did not. For the practice test group, the correlations between students' trait anxiety and performance on the practice test (r = -.079, p=.555) was trivial, as were correlations between performance on the first assessed test and students' state anxiety (r =-.160 , p=.186 ) and trait anxiety (r=-.203, p=.130). Similarly, there were no relationships between state anxiety and performance on the assessed test (r =-.191, p=.058) or between trait anxiety and performance on the assessed test (r=-.077, p=.622) for those students who did not sit the practice test. Again the scatter plots (see Figures 8 to 12) eliminate the possibility that the negligible linear correlation might be due to a curvilinear relationship.

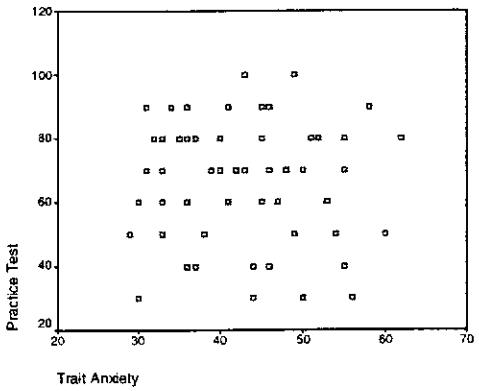


Figure 8. Practice test mark and trait anxiety for practice test group (Study 5)

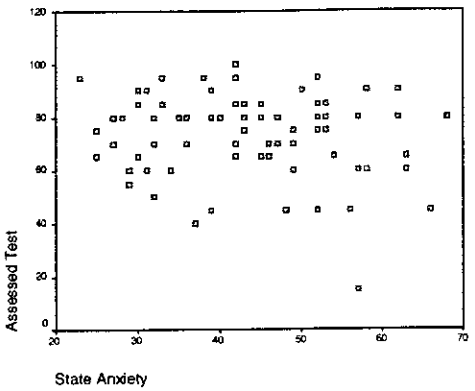


Figure 9. Assessed test mark and state anxiety for practice test group (Study 5)

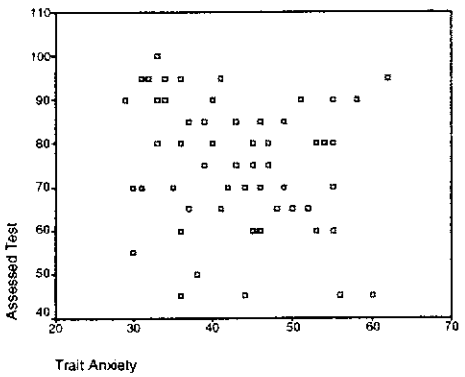


Figure 10. Assessed test mark and trait anxiety for practice test group (Study 5)

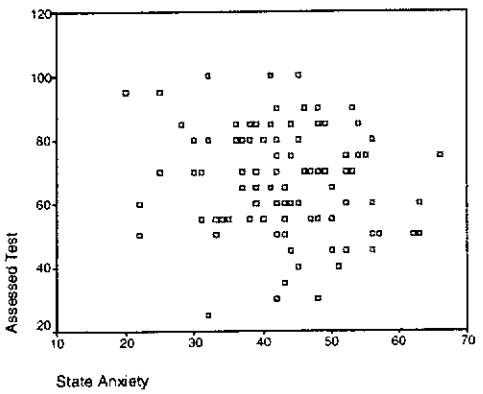


Figure 11. Assessed test mark and state anxiety for non practice test group (Study 5)

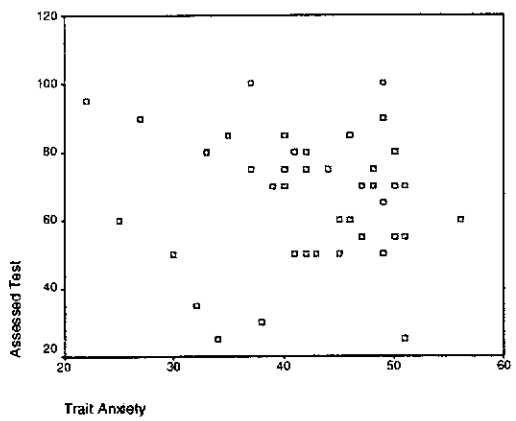


Figure 12. Assessed test mark and trait anxiety for non practice test group (Study 5)

Summary: Anxiety Component of Studies 3 and 5

Two groups of first year students from different subject disciplines produced similar results. No relationship was found between either trait or state anxiety and performance for either group. In Study 5, where the practice test was optional, no relationship was found between the anxiety measures and student choice to sit the practice test. A decision was made at this stage not to investigate anxiety any further in relation to performance on CML tests.

## Section 2 – Sex of Student

In Studies 2, 3, 5, 9 and 10, data about students' sex were collected from various questionnaires and used to examine whether the sex of the student was related to students' choice to sit a practice test, their performance on the assessed test or their anxiety levels.

### Choice to Sit the Practice test

Study 2

For this second semester group of Economics 101 students, 590 answered the question on sex on *Student Demographics - Study 2* (see Appendix E), 298 (50.5%) were male. Table 5.3 reports males and females choice to take the practice test. No statistically significant difference was found between males and females on choice to sit the practice test (chi-square =.143, p=.706).

Table 5.3 Relationship between Sex of Student and Decision to Take the Practice Test (Study 2)

| | Practice test use | |
|---|---|---|
| Sex of Student | Did first practice test | Did not do first practice test |
| Male | 227 | 70 |
| Female | 227 | 65 |

Study 9

The question asking students to report their sex was answered by 44 Medical Imaging 131 students, of whom 12 (27.3%) were male. The survey used was *Feedback Survey – Study 9* in Appendix I. Six males and 26 females did the first practice test, however the numbers are too small to use Pearson's Chi-Square (p=.038, one cell has an expected count of less than 5 and the minimum expected count is 3.27). Table 5.4 reports practice test use for males and females.

Table 5.4 Relationship between Sex of Student and Number of Practice Tests Taken (Study 9)

| | Practice test use | | | |
|---|---|---|---|---|
| Sex of Student | Did P1 & P2 | Did P1 or P2 | Did not do practice | *n* |
| Male | 2 | 4 | 6 | 12 |
| Female | 13 | 13 | 6 | 32 |
| Total | 15 | 17 | 12 | 44 |

Note. P1 = practice test 1, P2 = practice test 2

Study 10

*Feedback Survey – Study 10* (see Appendix J) was used with Accounting 100 students. Their sex was reported by 636 students, of whom 304 (47.8%) were male. Table 5.5 reports students' decision to take a practice test by sex. A statistically significant relationship was found between sex and student choice to sit two, one or no practice tests (chi-square=14.08, p=.001). Table 5.5 shows that a greater than expected number of female students sat two practice tests while approximately equal numbers of males and females sat either one or no practice test.

Table 5.5 Relationship between Sex of Student and Number of Practice Tests Taken (Study 10)

| Sex of Student | Practice test use | | | *n* |
|---|---|---|---|---|
| | Did P1 & P2 | Did P1 or P2 | Did not do practice test | |
| Male | 34 | 80 | 190 | 304 |
| Female | 74 | 81 | 177 | 332 |
| Total | 108 | 161 | 367 | 636 |

Note. P1 = practice test 1, P2 = practice test 2

## Performance on the Assessed Test

As it was apparent that those students who sat the practice test performed better on the assessed test than those who did not, an additional investigation examined the relationship between student's sex and test performance. This was done in two ways in order to minimise the possibility that if a relationship existed it applied only to the practice test group. For Studies 2 and 9 only the performance of the practice test group on the assessed test was considered but for Studies 3, 5 and 10, all students who sat the assessed test were included.

### Studies 2 and 9

Table 5.6 reports the results for the Economics 100 and Medical Imaging 131 practice test groups. For Economics students in Study 2 there was a statistically significant difference between the group mean marks obtained by males and females on the practice test (independent t=2.40, p=.017) with an effect size of .25. Males scored higher. When this practice test group was compared on their two assessed tests no statistically significant difference was found. The results also show no statistically significant difference between males and females on either assessed test for the Medical Imaging 131 students (Study 9).

Table 5.6 Relationship between Performance on the Assessed Test and Sex of Student for Students Who Sat the Practice Test (Studies 2 and 9)

| Discipline | Test | Sex | *n* | Mean | SD | *t* | *p* |
|---|---|---|---|---|---|---|---|
| Economics 100 | | | | | | | |
| (Study 2) | Practice 1 | Male | 182 | 64.15 | 14.90 | 2.40 | .017 |
| | | Female | 202 | 60.40 | 15.62 | | |
| | Assessed 1 | Male | 182 | 73.08 | 15.73 | .33 | .739 |
| | | Female | 202 | 72.55 | 15.20 | | |
| | Assessed 2 | Male | 182 | 67.69 | 15.80 | .56 | .576 |
| | | Female | 202 | 66.76 | 16.64 | | |
| Medical Imaging 131 | | | | | | | |
| (Study 9) | Assessed 1 | Male | 6 | 73.3 | 9.40 | .15 | .884 |
| | | Female | 26 | 74.04 | 10.77 | | |
| | Assessed 2 | Male | 2 | 60.00 | 4.24 | .23 | .821 |
| | | Female | 13 | 62.38 | 14.10 | | |

Studies 3, 5 and 10

Students in Psychology 113 (Study 3) and Accounting 100 (Study 5) reported their sex when responding to Spielberger's State-Trait Anxiety Inventory (see Appendix F) and Accounting 100 students (Study 10) responded to a similar question on *Feedback Survey - Study 10* (see Appendix J). Response rates to this question were 97.6%, 91.7% and 79.8% respectively.

Table 5.7 Relationship Between Performance on the Assessed Test or Optional Practice Test and Sex of Student (Studies 3, 5 and 10)

| Discipline | Test | Sex | *n* | Mean | SD | *t* | *p* |
|---|---|---|---|---|---|---|---|
| Psychology 113 (Study 3) | | | | | | | |
| | Assessed 1 | Male | 72 | 71.01 | 14.46 | -3.12 | .002 |
| | | Female | 295 | 76.12 | 11.96 | | |
| Accounting 100 (Study 5) | | | | | | | |
| | Practice 1 | Male | 30 | 66.00 | 20.78 | .31 | .756 |
| | | Female | 44 | 64.55 | 18.98 | | |
| | Assessed 1 | Male | 80 | 69.38 | 15.78 | .69 | .489 |
| | | Female | 96 | 67.60 | 17.70 | | |
| Accounting 100 (Study 10) | | | | | | | |
| | Practice 1 | Male | 81 | 65.06 | 19.05 | .91 | .364 |
| | | Female | 117 | 62.65 | 17.83 | | |
| | Assessed 1 | Male | 277 | 68.27 | 14.63 | 1.40 | .163 |
| | | Female | 319 | 66.54 | 15.49 | | |
| | Practice 2 | Male | 67 | 66.87 | 19.86 | .53 | .594 |
| | | Female | 112 | 65.18 | 20.84 | | |
| | Assessed 2 | Male | 281 | 63.99 | 18.31 | -1.69 | .092 |
| | | Female | 320 | 66.47 | 17.68 | | |

For the predominately female (80.4%) group of Psychology 113 students, the difference in performance between males and females was statistically significant (independent t=-3.12, p=.002) with an effect size of 0.41. Females scored higher. No statistically significant difference was found for either Accounting 100 group (Studies 5 and 10). Table 5.7 reports the results for the three groups.

## Anxiety Level

### Study 3

Statistically significant differences were found between students' reported levels of anxiety on three of the four measures for Psychology 113 students (Study 3). Females scored higher, which indicates greater anxiety, in all cases. Results are given in Table 5.8. However, as reported in the previous section, for the entire group there was no statistically significant correlation between anxiety level and performance on the CML tests.

Table 5.8 Relationship Between Sex of Student and Anxiety Levels (Study 3)

| Test | Sex | *n* | Mean | SD | *t* | *p* |
|---|---|---|---|---|---|---|
| State (pre practice test) | Male | 60 | 36.73 | 10.62 | 4.11 | .000 |
| | Female | 267 | 43.03 | 10.74 | | |
| Trait (pre practice test) | Male | 60 | 38.53 | 9.33 | 2.54 | .011 |
| | Female | 262 | 41.82 | 8.95 | | |
| State (pre assessed test) | Male | 66 | 40.71 | 11.83 | 3.63 | .000 |
| | Female | 255 | 46.60 | 11.75 | | |
| Trait (pre assessed test) | Male | 66 | 39.71 | 10.28 | .90 | .370 |
| | Female | 253 | 40.89 | 9.27 | | |

### Study 5

For the Accounting 100 students of Study 5, no statistically significant difference was found between males and females on either state or trait anxiety levels. Table 5.9 reports the results.

Table 5.9 Relationship Between Sex of Student and Anxiety Levels (Study 5)

| Test | Sex | *n* | Mean | SD | *t* | *p* |
|---|---|---|---|---|---|---|
| State | Male | 81 | 42.80 | 10.12 | 1.13 | .259 |
| | Female | 93 | 44.59 | 10.70 | | |
| Trait | Male | 31 | 45.10 | 8.06 | 1.18 | .241 |
| | Female | 61 | 42.93 | 8.42 | | |

Summary: Relationship Between Sex of Student and Use of CML

No conclusion can be drawn regarding the sex of the student and choice to sit a practice test. There was a statistically significantly relationship between sex and choice to sit the practice test for the Accounting students (Study 10) but not for the Economics students (Study 2). The Medical Imaging group (Study 9) was too small to allow reliable analysis. In most cases there were no performance differences by sex. However, in the predominately female Psychology 113 group (Study 3), females reported higher levels of anxiety on three of the four measures, and then also outperformed their male colleagues on the assessed test. It is of interest that there was no statistically significant difference between the males and females in this group on their trait anxiety before the assessed test. Spielberger (1966, 1972) saw trait anxiety as a relatively stable indicator of individual difference in anxiety susceptibility as well as an estimator of the probability that anxiety will be experienced in stressful situations.

## Section 3 - Student Response to CML and Use of the Practice Test

Results from surveys in this section are divided into three parts: first, students' general attitude to CML, second, students' use of the practice test and third, if and how students use the feedback supplied to incorrect responses. Details of the surveys were reported in Chapter 3 and copies are provided in Appendices D to M. In addition, student interviews are reported as these have potential to clarify and extend

information gained from the survey results reported in the three parts that form Section 3.

## Students' General Attitude to CML

This section reports survey results that consider students' general attitude to CML in an attempt to see how students use the CML system as well as to gauge students' opinions of the strengths and weaknesses in the CML system.

### Study 1

Two hundred and seventy seven Economics 101 students were given the survey *Attitude to CML – Study 1* (see Appendix D). Completed surveys were received from 213 (76.9%) students, of whom 118 (55.4%) were female. Students were not requested to give their name and were not followed up if they did not respond. Of the students responding to Question 1, 183 students (85.9%) said they were full time while 29 students (13.6%) said they were part time.

There were several potential reasons for students not handing in the forms. The student may not have sat the particular test, may have sat outside the allocated time frame and hence was missed by laboratory staff, did not hand back the survey, or was missed on a very busy day in the laboratory. However, the relatively high response rate suggests that the data are likely to be representative.

#### Students' Opinions about the Best Features of CML

Students were able to select any one or more of six options they believed were the best features of CML, and asked to add any additional information they wished. As Table 5.10 shows, most students selected "immediate feedback", followed by the freedom to select a convenient time as the most important features. About two thirds of students said they liked the removal of the concentrated pressure of a single mid-semester examination while 56% of students thought they should be offered the opportunity to resit a test if they wished to improve their mark. This resit

opportunity was already offered to students who failed an assessed test on the first attempt. Both the chance to take their own remedial action and the ability to revise on an ongoing basis were considered valuable by students.

Table 5.10 Which are the Best Features of the CML System? (Study 1)

| Option | *n* | Percent |
|---|---|---|
| Immediate feedback | 185 | 86.9 |
| Convenient time | 166 | 77.9 |
| Removes pressure | 139 | 65.3 |
| Resit opportunity | 119 | 55.9 |
| Remedial Action | 72 | 33.8 |
| Continuous revision | 70 | 32.9 |

Students were asked if they would like an optional practice test before each test, and 153 students (78.9%) said they would.

Student Choice for 20% of Mark

Table 5.11 reports student responses to the choice of assessment they wanted for twenty percent of the unit mark. Students were required to rank their preferences but were not obliged to rank all options supplied.

Table 5.11 Student Options for 20% of Mark (Study 1)

| Option | 1st Preference | Percentage | 2nd Preference | Percentage |
|---|---|---|---|---|
| CML Tests | 147 | 69.0 | 6 | 2.8 |
| Mid semester test | 9 | 4.2 | 38 | 17.8 |
| Another Essay | 10 | 4.7 | 22 | 10.3 |
| No preference | 19 | 8.9 | 4 | 1.9 |
| Other | 3 | 1.4 | 1 | 0.5 |

Students ranked CML tests well ahead of other options – a mid-semester test, a second essay (the unit already had one essay), no preference, or "other". This

question was on the second side of a double-sided sheet which 19 students failed to complete. Fifty-nine students left this question blank which may indicate that they did not complete page 2 of the survey or they intentionally left the question blank to register either a negative preference or no preference.

The results of this survey show that students enrolled in first year Economics were in favour of the CML package, believed that the practice test facility was helpful and would like an optional practice test before each assessed test. The data show that students are overwhelming in favor of retaining the CML tests in preference to a mid-semester test, an examination or any other option for this 20% of their mark.

Study 5 and Study 6

Accounting 100 students and Psychology 114 students completed the same three questions at the beginning of their respective surveys, *CML Usage – Study 5* in Appendix G and *CML Usage – Study 6* in Appendix H. These questions asked how much the student liked CML as a form of assessment, how well they felt the CML assessment measured their understanding of the unit and how difficult the system was to use. Surveys were returned by 75 students (39.5%) from the Accounting 100 group and 199 students (55.3%) from Psychology 114.

Both groups rated the computer managed learning system highly as a form of assessment, as shown in Table 5.12. More than half of the students in both groups selected 4 or 5 on the five-point scale. Very few responded that they hated the system. The results for Question 2 show that students from both groups believed the CML system was able to measure their understanding of the unit well. For Question 3, only two students, both from Psychology, rated the system as very difficult to use while 76% of Accounting students and 94% of Psychology students said it was not difficult.

Table 5.12 Students' Attitude to the CML System (Study 5 and Study 6)

| Questions | Study | Response (%) | | | | |
|---|---|---|---|---|---|---|
| | | 1[a] | 2 | 3 | 4 | 5[b] |
| **Q1**. How much do you like CML as a form of assessment? | 5 | 2.7 | 14.7 | 29.3 | 34.7 | 18.7 |
| | 6 | 7.0 | 11.6 | 27.1 | 28.1 | 26.1 |
| **Q2**. How well do you think CML assessment measures your own understanding of this unit? | 5 | 2.7 | 17.3 | 33.3 | 34.7 | 12.0 |
| | 6 | 15.6 | 23.1 | 26.1 | 26.6 | 8.5 |
| **Q3**. How difficult was CML to use? | 5 | 44.0 | 32.0 | 17.3 | 6.7 | 0.0 |
| | 6 | 78.4 | 15.6 | 3.3 | 1.5 | 1.0 |

Note. [a] is "*hate it*", "*not very difficult*" and "*not at all difficult*", respectively, for Q1, Q2 and Q3. [b] is "*like it a lot*", "*very well*" and "*very difficult*".

Summary: Attitude to CML component of Studies 1, 5 and 6

Students from Study 1 rated the immediate feedback offered by the CML system as well as the convenience of selecting their own testing time as important features of the testing system. When offered other choices for 20% of their mark, they were strongly in favour of retaining the CML tests. In both Study 5 and Study 6, approximately 50% of students rated the CML system highly with only a very small proportion expressing dislike. No students in Study 5 found the system difficult to use while only 1% of those in Study 6 had difficulty. Students from these groups are in favour of CML as a form of testing, do not find the system difficult to use and feel that it generally measures their understanding well.

## General Use of the Practice Test

The purpose of this section was to obtain additional information on students' use of the practice test.

### Study 5 and Study 6

A further two questions from the surveys: *CML Usage – Study 5* and *CML Usage – Study 6* are reported in Table 5.13. Students from Accounting 100 (Study 5) and Psychology 114 (Study 6) were asked how much study they did for the practice test and if they found the practice test a useful preparation for the assessed test. The table reports the results of the 49.3 % (37/75) of Accounting 100 survey respondents and the 47.2 % (94/199) of the Psychology 114 survey respondents who sat the practice test. Students reported that they did not study before the practice test, with 59.5% of Accounting students and 52.1% of Psychology students reporting no or little study and only 18.9% and 20.2%, respectively, claiming to have studied a substantial amount. However, most students reported that sitting the practice test "was helpful ... in preparing for the next CML test", with 62.1% and 71.2%, respectively, marking response 4 or 5.

Table 5.13 Responses from Practice Test Group Regarding Use of the Practice Test (Study 5 and Study 6)

| | | Response (%) | | | | |
|---|---|---|---|---|---|---|
| Questions | Study | 1 [a] | 2 | 3 | 4 | 5 [b] |
| **Q5.** How much did you study for the practice test? | 5 | 32.4 | 27.0 | 21.6 | 8.1 | 10.8 |
| | 6 | 26.6 | 25.5 | 27.7 | 18.1 | 2.1 |
| **Q6.** Do you think that the practice test was helpful to you in preparing for the next CML test? | 5 | 2.7 | 16.2 | 18.9 | 24.3 | 37.8 |
| | 6 | 8.5 | 6.4 | 13.8 | 37.2 | 34.0 |

Note. [a] is "*not at all*", [b] is "*a great deal/very helpful*"

Approximately half the students in each group reported not sitting the practice test. The most frequent reason given in each group was lack of time. This was reported by 11/38 (28.2%) Accounting students and 34/112 (30.4%) Psychology students. The next most frequent responses for the Accounting students included forgetting to sit the test (12/38, 30.8%) and not thinking the test would be helpful (6/38, 15.4%). Only one student gave previous use of the CML system as a reason. For the Psychology group, 32 students (28.6%), reported that they did not think the practice test would be helpful or that they had used CML previously (27/112, 24.1%). The fourth most common response was forgetting, with 10 students (8.9%) selecting this as an option. Some Psychology students gave more than one response so the percentages do not total 100%.

Students were asked to comment on the usefulness of the practice test as a preparation for the assessed test. Only 14 (37.8%) Accounting students wrote comments. Nine of these students reported that it gave them an idea of what the assessed test was like. Thirty-nine (54.9%) of the 71 Psychology students who wrote comments also reported that the practice test gave them an idea of what the assessed test was like and so what to prepare for. Five students (7%) said it showed them weak areas or areas to work on while two students made a general comment about the item type and standard. Five students (7%) said the practice test helped to make them less nervous about computers while seven students (9.8%) said it did not reflect the assessed test as it was either harder or easier.

General comments on the CML assessment were received from 125 Psychology students. Fifty-nine of the responses (47.2%) made a negative comment regarding either the items or the test. Eight comments were that the items tested facts rather than understanding, 13 said that the items did not show student knowledge, 19 comments related adversely to terminology commenting that "it was American". Three students said the testing did not help with examination preparation, while two students wanted items standardised as they felt some students received all easy items while others received more difficult items. Due to the predominately negative

comments by students, this question was repeated on a later survey with Psychology 114 students (Study 13). The Lecturer was sent an analysis of the testbank items and subsequently deleted some items. Those remaining were coded as easy or difficult. Students (Study 13) then received tests that selected half the items from those coded as easy and the remainder from the difficult ones.

Summary: Use of the Practice Test Component of Studies 5 and 6

Students in both Accounting 100 (Study 5) and Psychology 114 (Study 6) said they did little study before the practice test but approximately 70% of each group said they found the practice test helpful. It appears that the students in these two studies used the practice test to alert themselves to any of a number of issues including test and item format, level of difficulty, key areas of content and their own weak areas.

## Students' Use of Feedback

The main reason for surveys about the practice test was to find out how students use it because it provides feedback which can be used for formative assessment purposes. In Studies 11, 12 and 13 students were asked questions specific to their use of the feedback supplied on incorrect answers. Students responded to *Feedback Survey – Study 11* (see Appendix K), *Feedback Survey – Study 12* (see Appendix L) and *Feedback Survey – Study 13* (see Appendix M). A summary of responses is reported in Table 5.14. Question numbers used are those of *Feedback Survey – Study 11.*

Table 5.14 Summary of Student Responses to the Use of Feedback Supplied with Incorrect Answers (Studies 11, 12 and 13).

| Question | Response | Subject | | |
|---|---|---|---|---|
| | | Human Biology | Education | Psychology |
| **Q1.** After the practice test did you go through the items you answered incorrectly before you left the CML Lab? | Yes | 126 (94.0%) | 34 (97.1%) | 153 (89.0%) |
| | No | 8 (6.0%) | 1 (2.9%) | 19 (11.0%) |
| **Q2.** Was it useful to know the correct answer for a item you got wrong? | Yes | 133 (99.3%) | 34 (94.4%) | 153 (92.7%) |
| | No | 1 (0.7%) | 2 (5.6%) | 12 (7.3%) |
| **Q3.** If you were only told that your response was incorrect would you try to find the correct answer for yourself? | Yes | 72 (55.0%) | 35 (71.4%) | 213 (63.6%) |
| | No | 59 (45.0%) | 14 (28.6%) | 122 (36.4%) |
| **Q4.** If there was a textbook in the CML Lab for you to check correct answers for any item you got wrong would you use it before you left the CML Lab? | Yes | 73 (55.7%) | 34 (69.4%) | 200 (60.6%) |
| | No | 58 (44.3%) | 15 (30.6%) | 130 (39.4%) |
| **Q5.** Do you think it would be better ONLY to know that your answer was wrong so that you had to FIND the correct answer yourself? | Yes | 12 (9.0%) | 5 (10.2%) | 39 (11.5%) |
| | No | 121 (91.0%) | 44 (89.8%) | 299 (88.5%) |

Note. Some students omitted some questions

Study 11, 12 and 13

Surveys were returned by 136 (88%) of 155 students from Human Biology 134 (Study 11), 50 (77%) of the 65 Education 102 students (Study 12) and 340 (94%) of the 360 Psychology 114 students (Study 13). For both Questions 1 and 2, the

percentage of students who reported using the incorrect answer information was very high. For Question 1 percentages ranged from 89% to 97% and for Question 2, regarding the usefulness of being supplied with the correct answer, percentages ranged from 93% to 99%. Across the three studies, 62% of students said they would try to find the correct answer if it was not supplied (Q3). Between 56% and 69% of students said they would use a textbook supplied through the CML Laboratory to find the correct response (Q4). Only 56 students in total said that they favoured the option of not receiving the correct answer to a item they answered incorrectly, with percentages across the groups ranging from 9% to 11%.

Use of incorrect answer information

Table 5.15 reports students' comments to open-ended questions on the surveys completed in Studies 11, 12 and 13. Multiple responses were allowed so the percentages do not total 100%. Percentages are based on total responses not on number of students responding. Only those categories where the percentage of responses is greater than 5% are reported. Of the 134 responses received from 119 Human Biology 134 students (Study 11) regarding the use of the information supplied on incorrect answer, 88.0% of responses referred to the use of the information as a guide to error areas, key (content) areas or for revision of particular items. A further 12.6% referred to general revision. For Education 102 students, (Study 12) 60.9% selected the identification of error areas as important and another 30% referred to the use of the information for either item specific or general revision. The pattern for Psychology 114 students (Study 13) was rather different. The most common response was that the information acted as a motivator, a factor not mentioned by the other groups. Other common responses referred to use in alerting to error areas and issues specific to the particular item. Less than 40% of students from Studies 11, 12 and 13 added additional comment when given the opportunity (Q8). A small percentage of students from each group said they wanted to be given the correct answer to an incorrect response possibly to strengthen their initial response to Question 2 in Table 5.14. In addition, some students requested an explanation of why an answer was incorrect. This factor was more important to

Psychology students (35%) than to either Human Biology (12%) or Education students (11%). Approximately 30% of students from each group took the opportunity to make some positive comment about the CML system and the testing procedures.

Table 5.15 Student Comments Related to the Use of CML Information on Incorrect Responses across Studies 11, 12 and 13

| **Q7.** *How do you use the information about incorrect answers that you get from the CML system to help your preparation for the next CML test?* | | | |
|---|---|---|---|
| Comment | Human Biology *n*=119 | Education *n*=46 | Psychology *n*=296 |
| Identifies error areas | 58 (48.7%) | 28 (60.9%) | 86 (29.1%) |
| Identifies key content areas | 20 (16.8%) | | 19 (6.4%) |
| Item specific revision | 28 (23.5%) | 9 (19.6%) | 95 (32.1%) |
| General revision | 15 (12.6%) | 5 (10.9%) | 12 (4.1%) |
| Hard to remember | 7 (5.9%) | 3 (6.5%) | 24 (8.1%) |
| Do not use | | | 22 (7.4%) |
| Acts as a motivator | | | 135 (45.6%) |
| Identifies how item asked | | | 26 (8.8%) |
| **Q8.** *Any comments you can make regarding the usefulness, or otherwise, of the information that the CML system provides when it marks your answers would be helpful* | | | |
| Comment | Human Biology n=40 | Education n=18 | Psychology n=88 |
| Want to be given the correct answer | 9 (22.5%) | 2 (11.1%) | 16 (18.2%) |
| Want an explanation | 5 (12.5%) | 2 (11.1%) | 31 (35.2%) |
| Positive comment about the CML system | 14 (35.0%) | 7 (38.9%) | 26 (29.5%) |

Note. Multiple responses were allowed. Percent represents percentage of responses.

Summary of Survey Data for Study 11, 12 and 13

There is agreement across these three studies from different subject disciplines even though students' use of the CML system was different. Students generally look over their incorrect answers before leaving the CML Laboratory. Over half the students would find the correct answer if it was not supplied and approximately the same percentage would check a textbook supplied through the CML Laboratory to find the answer to an item they answered incorrectly. Students are in favour of being supplied the correct answer, with 11.5% being the highest percentage of students from any of the three studies feeling that it would be better to have to find it for themselves. The main uses of the feedback information given on incorrect answers include identification of error areas, key content areas and as a general revision tool. Students are positive about the CML system and many commented so even when not specifically requested to do so.

## Student Interviews

A total of 26 students from Studies 9, 11, 12 and 14 were interviewed regarding their response to CML and use of the feedback it supplied. Medical Imaging 131 students (Study 9) participated in longer interviews conducted outside the CML Laboratory. Students from Human Biology 134 (Study 11), Education 102 (Study 12) and Medical Imaging 132 (Study 14) responded to questions usually asked of them when they collected their error summary at the completion of a CML test. Table 3.4 (see p. 83) reports the number of students involved from each of these studies.

The main questions asked for information relating to use of the practice test, students' perceptions of the optimal conditions for a practice test, as well as asking students how they liked the particular testing situation used by their specific unit. Students in these studies used the CML system in a variety of ways. The Medical Imaging 131 and 132 students (Study 9 and Study 14) had an optional practice test before each of two assessed tests but the second practice test of Study 9 and the first practice test of Study 14 did not supply students with the correct answer to an

incorrect response. Students in Human Biology 134 (Study 11) had one practice test and were allowed two tries at each of their six CML assessed tests with the higher mark contributing towards their unit assessment. Tests did not have to be taken in order. Education 102 students (Study 12) had a similar test structure. They were allowed two tries at each of their five CML assessed tests and again counted the higher mark, but students were allowed multiple attempts at the practice test that preceded each assessed test.

The main questions addressed are listed in Chapter 3. Not all students responded specifically to each question and real names are not used in reporting student responses.

Student Responses Relating to the Practice Test Use

Anne (Study 9) didn't sit the first practice test as she had a car accident and "the time passed". She also failed to sit the second practice test as she "didn't get around to it" but, in retrospect, wishes she had sat the tests as a friend who did thought it helped. Catherine (Study 9) said that she had sat a practice test "to get a rough idea of the sections of the test" but she had not really taken it seriously because it was "just to see specifically what sections to study for the real test". She revised all sections between the practice test and the real test as "the practice test is not about getting it right...it's about narrowing down specific sections to study." Catherine (Study 9) also said it would be very helpful if she knew which sections of her course the modules that were reported on the error summary referred to, because "then you'd know where the majority of questions are coming from and therefore specific modules to study in more detail – and also the types of format of questions asked from specific modules – for example specific or general or ... questions about positioning, etc." Catherine was using the practice test as a formative assessment tool predominately to alert her to the type of items and the content areas they would be selected from. She was prepared to restudy or revise before the assessed test.

Mary (Study 9) sat both practice tests and was pleased she had. She wrote down the items she had made errors in when she left the CML laboratory and used this information to identify error and important areas. She used the practice test to gauge real test difficulty, item type, format etc. "There are lots of angles and we could have been asked lots about them but we were asked about joints". She was referring to the usefulness of the practice test in alerting her to the amount of detail/difficulty required in the assessed test. Her comments regarding the error summary were that it was useful to know which sections corresponded with textbook chapters but it definitely was not essential because you realise the areas you made mistakes in. David from Study 12, who had the option to sit the practice test more than once said he decided to go back and sit the practice test again " to get an idea of areas to work on". These students were using the practice test as a formative assessment tool and were prepared to restudy or revise prior to the assessed test.

Student Responses Relating to the Practice Test Characteristics

Students from Studies 9, 11, 12 and 14 were in agreement that the practice test and the assessed test should be the same length. One student commented specifically that as the practice test was only ten items, there was the possibility that on the assessed test you could be asked on areas that were not covered by the practice test due to the scarcity of items on it.

Student Responses Relating to the Number of Attempts at the Assessed Test

Students from all studies wanted two attempts at the assessed test with the higher mark to count towards their assessment. They prefer this to one practice test followed by one attempt at the assessed test. "I would study much more for the first [assessed] test because a practice test isn't seen and the mark attained has no consequences on what you actually get [as a grade], whereas a first test could be the higher of the two and you'd want to do at least all right just in case you completely stuff test 2" (Catherine – Study 9). A student from Study 11 described the idea of two attempts at a test as "excellent". She went further to say "I couldn't believe it

when the lecturer said we could have two attempts at a test and count the higher mark". This student said that if she scored more than 90% on her first try at the test she wouldn't repeat it. Other students said 50% or 60% would make them repeat the test while a higher score on the first try would be acceptable and they would not attempt the test the second time.

Three students (Study 11) who were taking their second try at the first assessed test were asked questions relating to the amount of study they did after the first try and before this second attempt. As a group they responded that they did little between attempts, with one student volunteering that she had scored 75% on try one and try two was "worth a second go as the higher mark is counted". Another student attempting the first assessed test for the second time commented that the two tries system was "fantastic", while an older student commented that "two tries gives you a sense of security". This last student planned to sit a first attempt at each test and then before the end-of-year examinations take the second attempt as a form of revision.

Students use the two-tries system in different ways. Two students who were both taking a second attempt at the first assessed test described how they approached the test. One had studied for the first try and did not revise between it and the second attempt while the other student had not studied for attempt one but revised before attempt two. One student (Study 11) who was taking attempt one at the first assessed test admitted to not studying for the test but was merely using it to alert himself to important areas. A second Study 11 student had prepared for this first attempt at the second assessed test and had also prepared for each of his two attempts at the first assessed test.

John (Study 11) admitted that the amount of study he did was "not as much as if I'd only had one go". Jenny and Sarah (Study 9) said that they did approximately half as much work for the practice test as they did for the assessed test.

## Summary of Interview Data

The interview data are in general agreement with the responses to student surveys reported previously. Students find a practice test useful, use it to alert themselves to their own error areas as well as important content areas. They prefer it to be of a similar length to the assessed test and to be supplied with the correct answers to an incorrect response. In the majority of cases they study little before the practice test.

When given the option of two attempts at an assessed test, the students' approach varies. Some students do more study than others, some take it seriously and some just treat it as a trial run. When the options are combined and students have both practice tests and two tries at an assessed test they again give a variety of responses. One interesting issue here was the order of testing for these students. Some students have an initial attempt at the assessed test and then re-evaluate their position, if required. This re-evaluation may mean attempting the practice test before taking the second attempt at the assessed test.

The fact that students are re-evaluating their position prior to the assessed test means that they are using the practice test (or the first attempt at a two-try assessed test) as a formative tool. They are looking for information from the practice test that will alert them to probabilities for the assessed test but they are going one step further and revising or restudying in order to improve their subsequent performance.

# CHAPTER 6

# SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

This thesis reports the results of studies set up to investigate formative assessment in the context of a computer managed learning (CML) practice test. The studies sought to determine whether taking the practice test affects performance on later CML assessed tests for first year university students and to determine the characteristics of the most effective CML practice test. The research took place in a large, centrally run testing laboratory at Curtin University of Technology which was being used predominately for summative assessment. Data collection occurred using the testing format dictated by the lecturer in the units concerned and a picture of the use of the CML system and the effects of the practice test was built up across a series of studies.

The general research question asked whether the practice test affects later CML performance and what characteristics make it most effective. Specifically, the study investigated the following questions:

1. How does the practice test affect performance on later computer managed learning (CML) assessed tests for first year university students?

   1(a) What is the effect of the amount of content overlap between the practice test and the assessed test?

   1(b) What is the effect of the length (number of items asked) of the practice test and what length do students prefer?

   1(c) What is the effect of feedback on incorrect responses and how do students use the feedback?

2. What are the characteristics of the most effective CML practice test?

A number of other factors that could have contributed to any improvement in performance following the practice test were investigated. These factors were subject discipline, ability, familiarity with the CML system, students' anxiety level, their sex, and their response to CML. The last factor includes students' attitude to CML generally and their use of the practice test.

The study focused on first year students because this was the area of greatest use of the CML Laboratory. Data were collected in a real testing environment through a series of small studies. Fourteen groups of students participated, with two groups participating by survey and interview only. In total there were fourteen practice tests and their associated assessed tests, six self-report questionnaires and ten student surveys. Some students were also asked to participate by way of interview. Data were obtained from these various sources and the findings pooled in an attempt to provide answers to the research questions.

## Findings

### Effect of the Practice Test on Performance on Later CML Assessed Tests

In this research, all studies where the practice test was investigated, except Study 7 (that is, Studies 1, 2, 3, 4, 5, 6, 8, 9, 10, 13, 14), have shown an improvement in mean student mark on the assessed test. That is, across various subject disciplines, there has been an improvement in mean student mark from the CML practice test to the CML assessed test. Further, those students who sat the practice test outperformed the group who did not, that is, students did better if they had done the practice test. . In Study 7 students improved their performance on that component of the assessed test that was covered by the practice test.

## Content Overlap between the Practice Test and the Assessed Test

The effect of content overlap between the practice and assessed tests was investigated in two ways. Firstly, quantitative data were obtained directly from CML tests. Studies were divided into those that covered 100% of the content of the assessed test and those that covered only a fraction of it. Secondly, additional information was gained from students via surveys and interviews.

Results of the nine studies (Studies 1, 2, 4, 6, 9(a), 9(b), 10(a), 13 and 14), where the practice test covered the same content as the respective assessed tests, showed improved student performance on the assessed test with effect sizes ranging from 0.45 to 2.06 (see Table 4.9). The average effect size was 1.14. In all cases, the difference between the test marks was statistically significant. The practice test group also performed better than the non practice test group on the assessed test. The difference achieved statistical significance for all except two groups (Study 4 and Study 9(a)).

Except for Study 7 students, students also improved their performance from the practice test to the entire assessed test when the practice test did not cover the entire content of the assessed test (Studies 3, 5, 10(b)). The effect sizes ranged from 0.17 to 0.41 (see Table 4.15) with an average effect size of 0.26. The average effect size for the difference in mean mark between Part A (content covered by the practice test) of the assessed test and Part B (new content) was 0.56 for the practice test group and 0.45 for those students who did not sit the practice test. That is, both groups performed better on Part A than on Part B. As these differences between Part A and Part B exist for both groups, the amount of content covered by the practice test can not explain all of the effect. Part B material was covered later in the semester so students had less time to revise. It may also have been more difficult or relied on consolidation of facts not thoroughly learnt from previous sections. These explanations are also consistent with the results from Study 7, where students did much better on Part A than Part B (see Table 4.12).

An additional example of how the benefits of practice tests can not be explained simply by considering the amount of content covered, comes from the one group of students (Study 8) whose practice test covered completely different content to that examined by the assessed test. Here the practice test group also performed better than the non practice test group.

It has been shown in all studies, except Study 7, that students sitting the practice test showed improved performance on the assessed test regardless of the amount of content covered by the practice test. However, students preferred the practice test to reflect the assessed test, indicating by responses to both surveys and interviews that the feedback provided by the CML system allowed them to identify error areas as well as important content areas. So, making the content coverage of both tests equivalent appears to be to students' advantage.

Comparison of the Length of the Practice and Assessed Tests

The length, or number of items, of the practice test was compared to the length of the assessed test to investigate whether the length of the practice test influenced student performance on the assessed test. Both quantitative and qualitative methods were used to determine an optimal length for a practice test. This was examined, firstly, by comparing results for those studies where the practice test and the assessed test covered the same content range and, secondly, by asking students to indicate the length of the practice test they thought was most desirable.

It was found that students showed improved performance from the practice test to the assessed test, despite tests of varying lengths, with effect sizes ranging from 0.45 to 2.06 (Studies 1, 2, 4, 6, 9(a), 9(b), 10(a), 13 and 14, see Table 4.16). For three studies (1, 2 and 14) where the practice test was the same length as the assessed test, the effect sizes ranged from 0.68 to 2.06. For the two studies (4 and 9(b)) where the practice test had the smallest proportion of items (1/3) effect sizes were 1.28 and 2.05.

The average effect size for the difference in mean mark on the assessed test between those students who sat the practice test and those who did not was 0.52 when the practice and assessed tests were the same length. When the test was half as long the average effect size was 0.3 while when it was less than half the length the effect size was 0.52. Hence, the number of items on the practice test does not appear to have a consistent relationship with student performance on the assessed test.

When asked in *Feedback Survey - Study 13*, (see Appendix M) students said they preferred to have a practice test that mirrored the assessed test in length. The reasons students gave added little information, because the most frequent reason given was that the practice test would replicate the real situation. Other reasons for wanting a practice test that was the same length or longer than the assessed test included the ability to provide a better sample of items. In addition, students referred to the weighting of individual items, that is, as the test length increased the individual item weighting decreased. This last reason regarding item weighting adds support to the practice test containing a similar number of items to the assessed test because students get a better estimation of time required for the test and their likely level of performance.

## Feedback from the Practice Test

The issue of feedback was addressed in two ways. Three groups of students in the Medical Sciences area (Studies 4, 9(b) and 14) received a practice test that informed them whether the response was correct or incorrect but did not supply the correct answer. Students from one of the original groups (Study 9) plus four additional groups (Studies 10, 11, 12 and 13) completed surveys which were used to gauge student opinion about the withholding of the correct response.

When students were told that a response was correct or incorrect but not supplied with the correct answer, they continued to show improved performance from the practice test to the assessed test. The effect sizes, for the difference between the

practice test and assessed test group mean mark ranged from 1.28 to 2.06, in these three studies (4, 9(b) and 14), however, the groups tested were quite small, ranging in size from 19 to 29 students. When the correct response was given (Studies 1, 2, 6, 9(a), 10(a) and 13) effect sizes were smaller, ranging from 0.45 to 1.03 (see Table 4.18), suggesting that there may be a benefit in supplying only confirmation that a response was correct or incorrect, that is, withholding the correct answer. However, the effect sizes are very variable in each condition.

When students were asked by survey or interview for their preference, they were adamant that they wanted to be given the correct answer. The uses they said they made of the feedback provided by the CML system can be divided into three categories: error detection, test clues and motivating factors. While students indicated that they used the feedback to correct errors on individual items, there was a stronger indication that the feedback was used to identify key content areas and important areas of personal weakness. Although students were strongly in favour of receiving the correct answer, it appears that they were using the feedback to alert themselves to areas to study rather than answers to individual items. The feedback may have been used as a tool to facilitate learning by providing the necessary incentive for students to put effort into revising prior to the assessed test.

### Additional Factors Having the Potential to Influence Results

In addition to the content covered by the practice test, its length and the type of feedback it supplied, additional factors were investigated. Subject discipline, ability and familiarity with the CML system were investigated as factors having the potential to influence students' choice to sit a practice test or students' performance on a CML test. In addition anxiety, sex of student and their response to CML, including the use of the practice test, were investigated as factors that may influence students' performance.

Subject Discipline

Previous quantitative reports in the CML literature have concentrated on a single subject discipline (Boyle et al., 1997; Charman & Elmes, 1998; Stanford & Cook, 1987). This thesis reported units from five subject disciplines, Economics, Psychology, Medical Sciences, Accounting and Education. In all subjects where the results from the practice and assessed tests were examined (Economics, Psychology, Medical Sciences and Accounting), those students who sat the CML practice test improved their performance on the assessed test and performed at a higher level on the assessed test than the group of students who did not sit the practice test. One small Psychology group (Study 7) showed improvement in the group mean mark from the practice test to Part A only of the assessed test. However, this does not contradict the conclusion that subject discipline did not affect student performance on these CML tests.

Ability

As this research progressed it became clear that those students who sat the practice test were performing better than those who did not. Possibly this could be explained by the more able students choosing to sit the practice test, so ability was investigated as a factor that may have contributed to student choice to sit the practice test. No link was found between ability as measured by university entry mark, final exam mark (excluding any CML component) in the current unit or exam mark in the prerequisite unit, and student choice to do an optional practice test.

Familiarity with the CML System

One factor which seemed to influence choice to sit the practice test was familiarity with the CML system. A positive impact, in terms of higher scores, has been seen across all studies where some students sat the practice test and some did not. Even, in the one unit where a practice test covered content that was not examined on the assessed test (Study 8), the group of students who sat this practice test outperformed their non practice test colleagues on the first assessed test. By the second assessed

test in this study, when all students had exposure to the CML system, the second assessed test, there was no statistically significant difference in performance between the groups. One possibility for the advantage seen initially for the practice test group is the benefit gained from using the CML system and seeing how the CML Laboratory functioned. The Study 8 students also had prior exposure to the multiple-choice Psychology items on their CML practice test even though these items were drawn from a different section of content than those used on the assessed test.

As well, familiarity with the CML system may contribute to students choosing not to sit an optional practice test. One student from Study 5 and 27 students from Study 6 gave previous CML use as a reason for not sitting the practice test. Students from Psychology 114 (Study 6) who had used the CML system in the previous semester had a lower participation rate on the optional practice test than groups from most other areas. Those units that offered students the option of a second practice test before their second assessed test also had a lower participation rate for the second practice test (Studies 9 and 10).

Student Anxiety

The research literature suggests that anxiety may have a detrimental impact upon student performance however this appeared not to be the case in the CML testing environment at Curtin University of Technology. The circumstances under which students sat CML tests may have minimised student anxiety. Students are able to choose their examination time, follow clearly defined procedures and receive procedural help from staff if required. Nevertheless, the possible effects of anxiety were examined by using the State-Trait Anxiety Inventory for Adults by Charles D. Spielberger (see Appendix F) in two studies. No relationship was found between either trait or state anxiety and performance on CML tests for either Psychology students (Study 3) or Accounting students (Study 5). Computer anxiety was not specifically measured, but students did not mention computer or test anxiety on any of the surveys when given the option to comment on the CML system. Contrary to

the view of Brosnan (1999) that computer anxiety is becoming increasingly prevalent, this study has found no evidence of any detrimental effect caused by student anxiety levels.

Anxiety was measured in Studies 5 and 3, but the practice test in Study 3 was compulsory, so only anxiety data from Study 5 were used to investigate students' choice to sit a practice test and their anxiety level. No relationship was found between the anxiety measures and student choice to sit the optional practice test.

Sex of Student

Five studies (2, 3, 5, 9 and 10) in this research investigated the possibility that sex may be a factor influencing either student choice to sit the practice test or performance on CML tests. Sex as a factor influencing choice to sit a practice test was investigated in Studies 2, 9 and 10. There was a statistically significantly relationship found for the Accounting students (Study 10) but this was not the case for the Economics students (Study 2). Analysis of the Medical Imaging group (Study 9) was unreliable due to the small group size. Sex as a factor affecting student performance on the assessed test was examined in two ways. Firstly, males' and females' performances on the assessed test were investigated using only the practice test group (Studies 2 and 9) and, secondly, in Studies 3, 5 and 10 the results of all students who sat the assessed test were included. A statistically significant relationship was found only in the predominately female Psychology 113 group (Study 3) that had a compulsory practice test. Females outperformed their male colleagues on the assessed test. So, for first year university students across Economics, Psychology, Medical Imaging and Accounting, sex as a factor affecting choice or performance has inconclusive results and needs further research.

Response to CML

Students were surveyed on their attitude to CML as this was considered to have potential to influence student choice to use the CML practice test. Results showed

that students were positive about the use of CML and, when given the option to select CML testing or another form of assessment for 20% of their mark, they were strongly in favour of retaining CML tests. They rated the immediate feedback offered by the CML system, as well as the convenience of selecting their own testing time, as important features of the testing system. As a group, they did not find the system difficult to use, and they felt that it measured their understanding effectively.

### Students' Use of the Practice Test and Feedback on Incorrect Responses

While admitting to doing little study before a practice test, the majority of students surveyed in Study 5 and Study 6 found the practice test helpful. They used it to alert themselves to test and item format, level of difficulty, key areas of content and their own weak areas. Students surveyed from Studies 11, 12 and 13 said they generally went through their incorrect answers before leaving the CML Laboratory. Over half the students said they would find the correct answer if it was not supplied, however, they were in favour of being supplied with the correct answer. These students identified similar ways of using information about incorrect answers as those students in Study 5 and Study 6 had attributed to the practice test, that is, identification of personal error and key content areas and as a general revision tool. Interview data confirmed the findings that students do little study before the practice test and prefer to be told the correct answer to a item they have answered incorrectly even when course notes are available for reference.

Based on the results of this research, the optimal use of this mainframe-based CML system as a formative assessment tool includes a practice test that is available before each CML assessed test. Student opinion adds additional detail, that is, they prefer that the practice test covers the same content and has the same number of items as the CML assessed test, provides the correct answer to an item answered incorrectly, and is able to be taken at a time determined by the student (within restrictions imposed by the lecturer).

## Discussion

Classroom formative assessment has been interpreted as including all activities that provide information which can be used as feedback to modify both teaching and learning activities (Black & Wiliam, 1998a). These teaching and learning activities must be interactive, because lecturers need to be aware of their students' progress in order to adapt their teaching to meet students' needs. In addition, students need to be aware of their position relative to the required knowledge or skills in order to modify their study when they have not reached the required level. Feedback on performance is specifically intended to improve and accelerate learning (Sadler, 1998). It has also been suggested that pretests act as advance organisers (Bangert-Drowns et al., 1991) and students in the studies in this research have nominated key functions of the practice test, and specifically the feedback supplied on incorrect answers, as triggers to help them identify areas to study. So, although the CML generated practice test used in this research has limited feedback features, it can function as a formative assessment tool which appears to facilitate student learning as shown by better performance on the assessed test by students doing a practice test compared with those who do not.

Consistent with papers based on the studies described in this thesis (Sly, 1999; Sly & Rennie, 1999a; Sly & Rennie, 1999b), a recently published article by Gretes and Green (2000) reported that those students who took computerised practice tests averaged higher grades than those who did not. Gretes and Green described results from two studies in which students were able to sit computerised practice tests one week prior to each of their two in-class tests and their final exam. The questions used were multiple choice and linked to course objectives, however, those items extracted from the item bank for use on practice tests were not used on in-class exams. On the practice tests, which were performed totally on the computer, students were able to skip items and then return to them before completing the test in a similar fashion to any paper and pen test. In addition, when students did submit an individual question for marking they were prompted for a second selection if the

initial one was incorrect. Apart from a total final score, rather than the item by item feedback given in the Curtin CML sytem, this second attempt on an initially incorrect response appears to be the only feedback students received. This is in contrast to studies reported in this thesis where students were either told the correct answer for an incorrect response or alternatively told that a response was incorrect as well as being given the opportunity to examine both their test paper and error summary before leaving the CML Laboratory. The differences in the approaches used for studies in this thesis and those reported by Gretes and Green (2000) together with their results demonstrate that the benefits of performing a practice test are not limited to the methods used in the CML Laboratory at Curtin University of Technology.

## Practice Test Parameters that May Influence Students' Performance

While it is apparent that the CML practice test is contributing to better performance, there appears to be no single factor to which this can be attributed. The content covered by the practice test, CML system familiarity, feedback and test length all were investigated as factors that may influence student performance.

### Content Covered by the Practice Test

As lecturers schedule practice tests in the weeks preceding the assessed test it is often the case that not all content that will be examined on the assessed test has been covered in the lectures. Practice tests that did not cover the entire content range of the assessed test always covered content that was on the first part of the assessed test. Within the constraints set by how CML is used at Curtin University of Technology, it has not been possible to find a group of students who have a practice test that reflects only the second part of their assessed test.

Students seem to find Part B (new content) of the assessed test difficult. Those who had new content examined on the assessed test, irrespective of having sat the practice test, did better on Part A (content covered by the practice test). It is possible

that items on content covered towards the end of a section are harder for all students as they may rely on early work that has not yet been consolidated and there has been less revision time. However, when there was zero overlap of content the group who did the practice test still performed better on the assessed test than the group who did not suggesting that there may be other contributing factors.

Familiarity with CML

Students who sat the practice test may have become familiar with the CML system, items or test type. It is possible that the beneficial effect seen with those students who sat the practice test was due to prior exposure to the CML system and that only one test is required for this purpose. However, this prior exposure may also be problematic. If lecturers believe that the beneficial effect of the practice test is only system familiarity they may not set a practice test before a second assessed test or in a new unit that has a prerequisite unit using the CML system. Students may choose not to sit a second practice test if they believe the benefit of the practice test is to familiarise themselves with a CML system they already know how to use. In this way, any gain from the use of the formative assessment available from the practice test may be missed.

Feedback

Students' use of feedback in the studies reported in this thesis appear to depend on the individual student. Some students use the feedback to predict the important sections on the assessed test while others use it to familiarise themselves with item characteristics such as format, degree of difficulty and amount of detail required. The fact that students consistently referred to the use of the feedback provided as being important in alerting them to their own errors as well as important content areas is a factor that must be considered when the content range of the practice test is determined. Some students named topics that were not covered by the practice test but subsequently examined on the assessed test. They felt the omission of these topics disadvantaged them.

In their recent article, Gretes and Green (2000) also ask if computerised practice tests affect student study habits or preparation time for assessed tests. This thesis provides some answers to this question. It has reported student survey and interview data that documents both the use students make of the feedback supplied to them by the practice test and how the practice test acts both as a motivator and to help them identify individual weak areas in their study.

Two things that became clear from the student surveys and interviews were that students use the correct answer to an incorrect response in a variety of ways and that they are keen to retain the option of receiving the correct answer. While CML test results show improvement when students are told only if a response is correct or incorrect, the students themselves believe they need to be informed of the correct answer. It appears that students are using the information gained from the feedback to identify and close the gap in their knowledge. This is consistent with the literature in other environments, which states that for feedback to make a contribution it must be used by students to identify areas where they do not reach the required standard and then used by students to help themselves achieve this standard (Black & Wiliam, 1998a; Ramaprasad, 1983; Sadler, 1989, 1998). Feedback is a key element because it provides information that the student can use to focus on areas of weakness. Even in this CML testing situation where feedback is limited, due to system constraints, results of formative assessment provided by the practice test along with comments from students have emphasised the importance of feedback.

It may also be that the students are using the external feedback they received from the error summary to monitor their performance, that is, generate their own internal feedback. This process has similarities to the operation of self-regulation identified in the literature (Butler & Winne, 1995) as it appears to allow students to monitor their progress in a manner that enables them to effect change. It also fulfills the criterion mentioned by Black and Wiliam (1998) that formative assessment is used as a diagnosis with the corresponding remedial action. However, students merely

state that they are using the supplied feedback to identify key areas of content for revision.

Some students referred to the motivating effect of the practice test. The literature also refers to the motivational effects and guidance of learning strategies that feedback can provide (Butler & Winne, 1995) as well as the information it can offer about both the study process and the mastery of the learning goals (Martens & Dochy, 1997). As students improved their performance from the practice test to the assessed test it may be that they used the information received on incorrect answers, test type or functioning of the CML system to change the way they studied.

Sadler (1989) states three points when explaining the importance of feedback in formative assessment. The learner has to (a) possess a concept of the standard, (b) compare the actual level of performance with the standard, and (c) take appropriate action which leads to closure of the gap. It appears that those students who sit the practice test are being provided with the information to address each of these points. They are made aware of the type and difficulty level of items, are able to compare the mark they obtain on the practice test with the mark their lecturer has informed them will be the pass mark on the assessed test, and have feedback from their practice test which they can use to direct their study for the assessed test. So while student and test characteristics varied across this research, the CML system, when used as a formative assessment tool, provided those students who elected to sit the practice test with the option to take whatever action was needed to close at least some of the gap between their initial level of performance and the standard required by their lecturer or the standard they set for themselves.

Length of the Practice Test

The number of items generated on a practice test was considered as potentially influencing results. The only conclusion that seems acceptable within the confines of the reported practice tests, ranging in length from one third of the assessed test to the

same length as the assessed test, is that, within reasonable limits, length is not a factor that affects performance on the assessed test. Students prefer to have both the practice and assessed tests of similar length, and suggest that they are using the practice test to gauge their performance against the assessed test, again signifying that the practice test is proving beneficial as a formative assessment tool. Students seem to be using the practice test as a trial for the real situation both on timing and depth of content coverage.

## Summary and Implications

The CML system has been shown to deliver many of the functions of formative assessment identified as important within the literature. Specifically, improved performance on the assessed test has followed the use of the practice test in all cases except Study 7 where the difference was not statistically significant. In all studies, student performance improved from the practice test to that part of the assessed test reflected by the practice test content. One group of students showed this improved performance when the practice test covered entirely different content from the assessed test suggesting that familiarity with the CML system may also be a contributing factor. Of course, as many students reported doing little study prior to the practice test, it is not surprising that their results showed improvement on the assessed test. Thus, the more compelling set of results is the consistently superior performance on the assessed test of the practice test group compared to the non practice test group in all those studies where the comparison was possible.

Students say they want the practice test to reflect the assessed test both in content coverage and length. In addition, students say they want feedback on incorrect responses. This feedback appears to be the major reason why the practice test is beneficial as students say they are using it to identify both important content areas and their own areas of weakness. So, in the eyes of students, the most effective practice test is one that is the same length as the assessed test, covers the same content range and gives students feedback on their responses to test items.

A clear conclusion from this research is that the CML system should be used for the provision of a practice test that serves only a formative purpose. So that students can make best use of the feedback opportunity the practice test provides, lecturers need to encourage student participation, not just on an initial practice test but on all practice tests offered. Also, students need to be encouraged to review their error summary with their test paper before they leave the CML Laboratory. In preparing instructions for the generation of the practice test lecturers should consider students' views that the optimal practice test covers the same content as the assessed test with the same number of items and provides the correct answer for an item answered incorrectly.

A related but important issue which was not specifically investigated in this research concerns the use the lecturers make of the feedback to them provided by the CML Laboratory. Lecturers receive two types of feedback, a report of student grades and an analysis of testbank items. The grades report allows lecturers to identify those students at risk early in the unit and hence provides lecturers with an opportunity to intervene. In addition, feedback on testbank items provides an overview of class performance which may influence the time lecturers allocate to difficult topics and the rate they progress through the unit. Thus, feedback provides information that can be used to examine the quality and performance of items, as well as the performance of students on various parts of the practice test.

There is little evidence to suggest that lecturers use the supplied feedback, however if they could be encouraged to do so there is potential for changes to teaching practice. For example, the quality of the item banks could be continuously revised and upgraded with consequent benefits to students. Perhaps more importantly, by reviewing and responding to students' performance on practice tests, lecturers could amend sequencing of content, vary time devoted to particular topics and tailor more effectively the materials covered in tutorials or given to students as resources. These kinds of changes would result in an enhanced teaching program. How to encourage lecturers to make more use of this feedback remains a problem, because they say

that with heavy teaching loads they find it easier to do what has been done previously.

The studies in this research have used a mainframe-based CML system operating in a secure testing environment where tests were computer generated, but paper-based, and feedback on responses has been limited to the provision of the correct response. As technology changes, it has become evident that the future for CML is in the Web-based environment. With new Web-based CML packages becoming available students will be able to sit and mark tests totally on the computer. Packages in this environment offer more sophisticated feedback than the one used in this research as many already include the provision for an explanation of why a response is correct or incorrect. These packages also allow lecturers easier access to item bank statistics than is possible on a mainframe-based system.

The Web-based environment is more flexible but less secure and student identification and software viruses are potential issues which need to be addressed. Further, there are resource implications which need to be investigated. For example, the time differences between students using Web-based systems compared to paper-based systems have potential to influence CML viability. With the existing system, students spend most of the testing time sitting at a desk with the test paper and the computer is used only to generate and mark the test. If the existing CML Laboratory were to run a Web-based CML system then each student would require a computer for the entire time taken for a test. Clearly this would create considerable equipment demands and within the present accommodation restrictions, support might need to be provided from other venues, such as on-campus open-access computer laboratories or by students using remote access. Again the issues of security and virus protection are raised.

Despite resource problems, the Web-based systems offer a number of advantages, including the enhanced feedback opportunities and the better access to item banks. It seems likely that the benefits of the practice test demonstrated in this research would

also be demonstrated in a Web-based testing environment, but this would need to be tested in new research. Other possible research topics include investigation of student use of the more detailed feedback possible. This could provide an avenue into a closer examination of self-regulated learning (Zimmerman, 1994) than was possible in this study. How lecturers use the system is another research topic of interest. Will lecturers use the opportunity provided by easier access to obtain and use feedback on particular items in their item banks? Will item banks be updated more frequently in this new environment? As lecturers become familiar with these new software packages, will they allow unsupervised practice tests for students? If so, do students use the practice tests in the same way, or do they develop different skills in self-monitoring of their performance? If formative tests become unsupervised will lecturers be prepared to have two item banks so that those items used for summative assessments are kept secure?

As pointed out by Thornton, Kevill, and Sly (2000), "In developing learning environments, which both support and promote the integration of life-long learning skills within their programs, tertiary institutions have to re-think and re-engineer the way they structure the students' learning experience" (p. 33). One way to develop these supportive learning environments is to move towards a more student-centered flexible environment that utilises technology effectively. While CML systems are changing rapidly, this research has shown that at least one CML system appears to have the potential to facilitate student learning by the provision of feedback and by system flexibility. This flexibility is both in the provision of tests in which the items can be randomised and the freedom for a student to select different testing times. It remains for future research to both confirm and extend the present findings, related to the beneficial effect of the practice test and especially the provision of feedback and flexibility, with other computer testing systems.

This study has provided a range of information about the use of the practice test in the CML environment that has not previously been reported. It has led to a more informed understanding of the use of CML as a tool for learning rather than only as

a summative testing technique. In particular, it has contributed to an improved understanding of students' use of the feedback supplied by this CML system. This feedback appears to be the key feature of the value of the practice test, because it enables students to test their current understanding and identify areas for further study. The superior performance on the assessed test by the students who had taken a practice test compared with those who did not, has been persuasive in promoting the use of the practice test at Curtin University of Technology.

# REFERENCES


Ablard, K. E., & Lipschultz, R. E. (1998). Self-regulated learning in high-achieving students: Relations to advanced reasoning, achievement goals, and gender. *Journal of Educational Psychology, 90*(1), 94-101.

Adams, G. S. (1964). *Measurement and evaluation in education, psychology, and guidance.* New York: Holt, Rinehart & Winston.

Aiken, L. R. (1987). Testing with multiple choice items. *Journal of Research and Development in Education, 20*, 44-58.

Alpert, R., & Haber, R. (1960). Anxiety in academic achievement situations. *Journal of Abnormal and Social Psychology, 61*, 207-215.

Ames, C. (1984). Achievement attributions and self-instructions under competitive and individualistic goal structures. *Journal of Educational Psychology, 76*, 478-487.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of educational Psychology, 84*(3), 261-271.

Ames, C., & Archer, J. (1987). Mothers' belief about the role of ability and effort in school learning. *Journal of Educational Psychology, 79*, 409-414.

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology, 80*, 260-267.

Anastasi, A. (1976). *Psychological testing* (4th ed.). New York: Macmillan.

Anderson, J. (1989). Sex-related differences on objective tests among undergraduates. *Educational Studies in Mathematics, 20*, 165-177.

Anderson, S. B., & Sauser, W. I. (1995). Measurement of test anxiety: An overview. In C. D. Spielberger & P. R. Vagg (Eds.), *Test anxiety: Theory, assessment and treatment* (pp. 15-33). Washington, DC: Taylor & Francis.

Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques*. San Francisco: Jossey-Bass Publishers.

Ball, S. (1995). Anxiety and test performance. In C. D. Spielberger & P. R. Vagg (Eds.), *Test anxiety: Theory, assessment, and treatment* (pp. 107-113). Washington, DC: Taylor & Francis.

Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*(2), 213-238.

Barnett-Foster, D., & Nagy, P. (1996). Undergraduate student response strategies to test questions of varying format. *Higher Education, 32*, 177-198.

Bedell, J. R., & Marlowe, H. A. (1995). An evaluation of the anxiety scales: Convergent, divergent, and predictive validity. In C. D. Spielberger & P. R. Vagg (Eds.), *Test anxiety: Theory, assessment, and treatment* (pp. 35-45). Washington, DC: Taylor & Francis.

Bennett, R. E., Rock, D. A., & Wang, M. (1991). Equivalence of free-response and multiple-choice items. *Journal of Eductional Measurement, 28*(1), 77-92.

Biggs, J. B. (1973). Study behavior and performance in objective and essay formats. *Australian Journal of Education, 17*, 157-167.

Biggs, J. B. (1978). Individual and group differences in study processes. *British Journal of Educational Psychology, 48*, 266-279.

Biggs, J. B. (1996). Assessing learning quality: Reconciling institutional, staff, and educational demands. *Assessment and Evaluation in Higher Education, 21*, 5-15.

Biggs, J. B., & Moore, P. J. (1993). *The process of learning*. Sydney, Australia: Prentice Hall.

Birenbaum, M., & Nasser, F. (1994). On the relationship between test anxiety and test performance. *Measurement and Evaluation in Counseling and Development, 27*, 293-301.

Birenbaum, M., & Tatsuoka, K. K. (1987). Open-ended versus multiple-choice response formats - It does make a difference for diagnostic purposes. *Applied Psychological Measurement, 11*, 385 - 395.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-74.

Black, P., & Wiliam, D. (1998b). *Inside the black box.* London: School of Education, King's College London.

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1, Cognitive domain.* London: Longman Group.

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction, 7*(2), 161-186.

Boyle, A. P., Bryon, D. N., & Paul, C. R. C. (1997). Computer-based learning and assessment: A palaeontological case study with outcomes and implications. *Computers and Geosciences, 23*(5), 573-580.

Bozionelos, N. (1996). Psychology of computer use: XXXIX. Prevalence of computer anxiety in British managers and professionals. *Psychological Reports, 78*, 995-1002.

Brosnan, M. (1999). Computer anxiety in students: Should computer-based assessment be used at all? In S. Brown, P. Race, & J. Bull (Eds.), *Computer-assisted assessment in higher education* (pp. 47-54). London: Kogan Page Limited.

Brown, S., & Knight, P. (1994). *Assessing learners in higher education.* London: Kogan Page.

Bull, J. (1999, June). Update on the national TLTP3 project: The implementation and evaluation of computer-assisted assessment. In M. Danson & R.

Sherrat (Eds.), *Proceedings of flexible learning: Third annual computer assisted assessment conference* (pp. 11-17). Loughborough, UK: Loughborough University.

Burden, F. R. (1992). Computer Managed Learning in action: Application to large classes. In B. Chia, R. Pennell, & R. Sims (Eds.), *Proceedings of the 1992 Conference of the Australian Society for Computers in Learning in Tertiary Education: A future promised* (pp. 41-44). Sydney, Australia: Australian Society for Computers in Learning in Tertiary Education.

Burke, M. E., Critchley, R. H., & Lessells, G. S. (1989). CML at the University of Limerick - the mathematics experience. *Conference proceedings for the first international CML users' conference: Consolidating the gains and planning future directions* (pp. PS4A). Limerick, Ireland: University of Limerick.

Butler, D., & Winne, P. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245-281.

Cameron, J., & Pierce, D. P. (1994). Reinforcement, reward, and intrinsic motivation: a meta-analysis. *Review of Educational Research, 64*, 363-423.

Cann, A. J., & Pawley, E. L. (1999). Automated online tutorials: new formats for assessment on the WWW. In S. Brown, P. Race, & J. Bull (Eds.), *Computer-assisted assessment in higher education* (pp. 40-45). London: Kogan Page.

Carlson, R. E., & Wright, D. G. (1993). Computer anxiety and communication apprehension: Relationship and introductory college course effects. *Journal of Educational Computing Research, 9*(3), 329-338.

Cartwright, N. (Ed.). (1997). *Assessment and feedback - A handbook for tertiary teachers.* Ballarat, Victoria: University of Ballarat.

Carver, C. S., & Scheier, M. F. (1981). *Attention and self-regulation: A control theory approach to human behavior.* New York: Springer-Verlag.

Carver, C. S., & Scheier, M. F. (1990). Origins and functions of positive and negative affect: A control-process view. *Psychological Review, 97*, 19-35.

C.B.T.S. (Australia) Pty Limited. (1995). *LMS Learning Management Systems*. (Vol. V3.5): Melbourne, Australia: Campus America, Inc.

Charman, D. (1999). Issues and impacts of using computer-based assessment (CBAs) for formative assessment. In S. Brown, P. Race, & J. Bull (Eds.), *Computer assisted assessment in higher education* (pp. 85-93). Birmingham: Kogan Page.

Charman, D., & Elmes, A. (1998). A computer-based formative assessment strategy for a basic statistics module in geography. *Journal of Geography in Higher Education, 22*(3), 381-385.

Clark, B. (1988). *Growing up gifted* (3rd ed.). Columbus, OH: Merrill Publishing Company.

Cohen, B. A., & Waugh, G. W. (1989). Assessing computer anxiety. *Psychological Reports, 65*, 735-738.

Coomb, V., & Cooper, K. (1992). Beyond the 1990's: A future promised through successful application of CBL & CML development techniques with the business sector. In B. Chia, R. Pennell, & R. Sims. (Eds.), *Proceedings of the 1992 Conference of the Australian Society for Computers in Learning in Tertiary Education: A future promised* (pp. 45-52). Sydney, Australia: Australian Society for Computers in Learning in Tertiary Education.

Corno, L. (1986). The metacognitive control components of self-regulated learning. *Contemporary Educational Psychology, 11*, 333-346.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*(4), 438-481.

Culler, R. E., & Holahan, C. J. (1980). Test anxiety and academic performance: The effects of study related behaviors. *Journal of Educational Psychology, 72*(1), 16-20.

Dalton, D. W., & Goodrum, D. A. (1991). The effects of computer based pretesting strategies on learning and continuing motivation. *Journal of Research on Computing in Education, 24*(2), 204-213.

Doig, S. M. (June 1999). Developing an understanding of the role of feedback in education. *Teaching and Education News, 9.*

Donnelly, A. (1989). Serving diverse student needs using educational technology. *Conference proceedings for the first international CML users' conference: Consolidating the gains and planning future directions* (pp. PS14). Limerick, Ireland: University of Limerick.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1*, 170-177.

Dweck, C. (1986). Motivational processes affecting learning. *American Psychologist, 41*, 1040-1048.

Dweck, C. S., & Elliot, E. S. (1983). Achievement motivation. In E. M. V. E. Hetherington (Ed.), *Handbook of Child Psychology: Socialization, Personality, and Social Development* (Vol. 4, pp. 643-691). New York: Wiley.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*(95), 256-273.

Ebel, R., & Frisbie, D. A. (1986). *Essentials of educational measurement.* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Entwistle, N. (1981). *Styles of learning and teaching.* Chichester, NY: Wiley.

Fleming, P. (1987). Computerised formative assessment - A powerful new tool. In J. Barrett, & J. Hedberg (Eds.), *Using computers intelligently in tertiary education. A collection of papers presented to Australian Society for Computers in Learning* (pp. 134-137). Sydney, Australia: Australian Society for Computers in Learning in Tertiary Education.

Folin, O., Demis, W., & Smillie, W. G. (1914). Some observations on emotional glycosuria in man. *Journal of Biological Chemistry, 17*, 519-520.

Gipps, C. (1994). Developments in educational assessment: What makes a good test? *Assessment in Education, 1*(3), 283-291.

Graham, S., & Golan, S. (1991). Motivational influences on cognitive: Task involvement, ego involvement, and depth of information processing. *Journal of Educational Psychology, 83*, 187-194.

Green, B. F. (1981). A primer of testing. *American Psychologist, 36*(10), 1001-1011.

Gretes, J. A. & Green, M. (2000). Improving undergraduate learning with computer-assisted assessment. *Journal of Research and Computing in Education,* 33(1), 46-54.

Gronlund, N. F. (1982). *Constructing achievement tests.* Englewood Cliffs, NJ: Prentice Hall Inc.

Gross, T. F., & Mastenbrook, M. (1980). Examination of the effects of state anxiety on problem-solving efficiency under high and low memory conditions. *Journal of Educational Psychology, 72*(5), 605-609.

Harrison, S. R., & Stanford, J. D. (1989). Some experiences in the use of CML for large-class teaching of statistical methods. In *Conference proceedings for the First international CML users' conference: Consolidating the gains and planning future directions* (pp. PS5). Limerick, Ireland: University of Limerick.

Hedges, L. (1981). Distributed theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107-128.

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research, 58*, 47-77.

Heywood, J. (1977). *Assessment in higher education.* New York: John Wiley & Sons.

Hill, J. R. (1981). *Measurement and evaluation in the classroom* (2nd ed.). Columbus, OH: Charles E. Merrill Publishing Company.

Honeyman, D. S., & White, W. J. (1987). Computer anxiety in educators learning to use the computer: A preliminary report. *Journal of Research on Computing in Education, Winter*, 129-138.

Hunsley, J. (1985). Test anxiety, academic performance, and cognitive appraisals. *Journal of Educational Psychology, 77*(6), 678-682.

Isaacs, G. (1994). *Multiple choice testing* (Vol. 16). Campbelltown, NSW: Higher Education Research and Development Society of Australasia.

Jay, T. J. (1981). Computerphobia: What to do about it. *Educational Technology* (January), 47-48.

Johnstone, R., Patterson, J., & Rubenstein, K. (1998). *Improving criteria and feedback*. Sydney: Cavendish Publishing Pty. Ltd.

Kelley, C. L., & Charness, N. (1995). Issues in training older adults to use computers. *Behaviour and Information Technology, 14*(2), 107-120.

Koohang, A. A. (1987). A study on the attitudes of pre-service teachers towards the use of computer. *Educational Communications and Technology Journal, 35*(3), 145-149.

Kulger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254-284.

Kulhavy, R. (1977). Feedback in written instruction. *Review of Educational Research, 47*(1), 211-232.

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review, 1*, 279-308.

Kulik, J., & Kulik, C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58*(1), 79-97.

Laurillard, D. (1996). *Rethinking university teaching*. New York: Routledge.

Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports, 20*, 975-978.

Lukhele, R., Thissen, D., & Wainer, H. (1994). On the relative value of multiple-choice, constructed response, and examinee-selected items in two achievement tests. *Journal of Education Measurement, 31*(3), 234-250.

Maki, R. H., & Serra, M. (1992). Role of practice tests in accuracy of test prediction on text material. *Journal of Educational Psychology, 84*(2), 200-210.

Mandler, G., & Sarason, S. B. (1952). A study of anxiety and learning. *Journal of Abnormal and Social Psychology, 47*, 166-173.

Mansen, T. J., & Haak, S. W. (1996). Evaluation of health assessment skils using a computer videodisc interactive program. *Journal of Nursing Education, 35*, 382-384.

Martens, R., & Dochy, F. (1997). Assessment and feedback as student support devices. *Studies in Educational Evaluation, 23*(3), 257-272.

Martin, E. (1999). *Changing academic work: Developing the learning university*. Buckingham, England: Society for Research into Higher Education and Open University Press.

Marton, F., & Saljo, R. (1976). On qualitative differences in learning: Outcome and process. *British Journal of Educational Psychology, 46*, 4-11.

Meece, J. L. (1994). The role of motivation in self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-Regulation of Learning and Performance* (pp. 25-44). Hillsdale, NJ: Lawrence Erlbaum Associates.

Morgenstern, C. F., & Renner, J. W. (1984). Measuring thinking with standardized test. *Journal of Reserch in Science Teaching, 21*, 639-648.

Mory, E. H. (1992). The use of informational feedback in instruction: Implications for future research. *Educational Technology Research and Development, 40*(3), 5-20.

Murphy, P. F. (1996). Assessment practices and gender in science. In L. H. Parker, L. J. Rennie, & B. J. Fraser (Eds.), *Gender, science and mathematics: Shortening the shadow*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Murphy, R. J. L. (1982). Sex differences in objective test performance. *British Journal of Educational Psychology*, 213-219.

Nicholls, J. G. (1984). Achievement motivation: Conception of ability, subjective experience, task choice, and performance. *Psychological Review, 91*, 328-346.

Nicholls, J. G., Cobb, P., Wood, T., Yackel, E., & Patashnick, M. (1990). Assessing students' theories of success in mathematics: Individual and classroom differences. *Journal for Research in Mathematics Education, 21*, 109-122.

Nicholls, J. G., Patashnick, M., & Nolen, S. B. (1985). Adolescents' theories of education. *Journal of Educational Psychology, 77*, 683-692.

Nightingale, P., Wiata, I. T., Toohey, S., Ryan, G., Hughes, C., & Magin, D. (1996). *Assessing Learning in Universities*. Syndey, Australia: University of New South Wales Press.

Nolen, S. B., & Haladyna, T. M. (1990). Motivation and studying in high school science. *Journal of Research on Science Teaching, 27*, 115-126.

Nolen, S. B. (1988). Reasons for studying: Motivational orientations and study strategies. *Cognition and Instruction, 5*, 269-287.

Okebukola, P. A. (1993). The gender factor in computer anxiety and interest among some Australian high school students. *Educational Research, 35*(2), 181-189.

Paris, S. G., & Oka, E. R. (1986). Self-regulated learning among exceptional children. *Exceptional Children, 52*, 103-108.

Parker, L. H., & Rennie, L. J. (1995). Equitable assessment strategies. In B. J. Fraser & K. Tobin (Eds.), *International handbook of science education.* (pp. 897-910). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Parsons, D., & Hunter, R. (1987). Computer managed learning in electronics. In J. Barrett, & J. Hedberg (Eds.), *Using computers intelligently in tertiary education. A collection of papers presented to Australian Society for Computers in*

*Learning* (pp. 345-351) Sydney, Australia: Australian Society for Computers in Learning in Tertiary Education.

Paterson, C. C. (1996). Self-regulated learning and academic achievement of senior biology students. *Australian Science Teachers Journal, 42*(2), 48-52.

Paulman, R. G., & Kennelly, K. J. (1984). Test anxiety and ineffective test taking: Different names, same construct. *Journal of Educational Psychology, 76*(2), 279-288.

Perkins, M. (1999). Validating formative and summative assessment. In S. Brown, P. Race, & J. Bull (Eds.), *Computer-assisted assessment in higher education* (pp. 55-62). Birmingham: Staff and Educational Development Association.

Phye, G. D. (1979). The processing of informative feedback about multiple-choice test performance. *Contemporary Educational Psychology, 4*, 381-394.

Pintrich, P., & De Groot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*, 33-40.

Pintrich, P. R., & Garcia, T. (1991). Student goal orientation and self-regulation in the college classroom. In M. L. Maehr & P. R. Pintrich (Eds.), *Advances in Motivation and Achievement* (Vol. 7, pp. 371-402). Greenwich, CT: JAI Press.

Piper, D. W., Nulty, D., & O'Grady, G. (1996). *Examination practices and procedures in Australian universities: Summary report.* Canberra: Australian Government Publishing Service.

Pope-Davis, D. B., & Twing, J. S. (1991). The effects of age, gender, and experience on measures of attitude regarding computers. *Computers in Human Behavior, 7*, 333-339.

Pritchett, N. (1999). Effective question design. In S. Brown, P. Race, & J. Bull (Eds.), *Computer-assisted assessment in higher education* (pp. 29-38). Birmingham: Staff and Educational Development Association.

Race, P. (1998). The art of assessing. URL: http://www.lgu.ac.uk/deliberations/assessment/artof.html, 2/3/1998.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science, 28*(4-13).

Ramsden, P. (1985). Student learning research: Retrospect and prospect. *Higher Education Research and Development, 4*, 51-69.

Rayner, S., & Riding, R. (1997). Towards a categorisation of cognitive styles and learning styles. *Educational Psychology, 17*(1 & 2), 5-27.

Riding, R. J. (1997). On the nature of cognitive style. *Educational Psychology, 17*, 29-49.

Riding, R. J., & Cheema, I. (1991). Cognitive styles - an overview and integration. *Educational Psychology, 11*, 193-215.

Rolfe, I., & McPherson, J. (1995). Formative assessment: How am I doing? *The Lancet, 345/8953*, 837-839.

Roper, W. J. (1977). Feedback in computer assisted instruction. *Programmed Learning and Educational Technology, 14*, 43-49.

Rosen, L., & Maquire, P. (1990). Myths and realities of computerphobia: a meta-analysis. *Anxiety Research, 3*, 175-191.

Rosen, L. D., Sears, D. C., & Weil, M. M. (1993). Treating technophobia: A longitudinal evaluation of the computerphobia reduction program. *Computers in Human Behaviour, 9*, 27-50.

Rowntree, D. (1987). *Assessing students: How shall we know them?* London: Kogan Page.

Sadler, D. R. (1989). Formative assessment and design of instruction systems. *Instructional Science, 18*, 119-144.

Sadler, D. R. (1998). Formative assessment: revisiting the territory. *Assessment in Education, 5*(1), 77-83.

Sadler-Smith, E. (1997). "Learning Style": Frameworks and instruments. *Educational Psychology, 17*(1 and 2), 51-63.

Salomon, G., & Globerson, T. (1987). Skill may not be enough: The role of mindfulness in learning and transfer. *International Journal of Educational Research, 11*, 623-637.

Sambell, K., Sambell, A., & Sexton, G. (1999). Student perceptions of learning benefits of computer-assisted assessment: a case study in electronic engineering. In S. Brown, P. Race, & J. Bull (Eds.), *Computer-assisted assessment in higher education* (pp. 179-191). London: Kogan Page.

Sarason, I. G. (1972). Experimental approaches to test anxiety: Attention and the uses of information. In C. D. Spielberger (Ed.), *Anxiety: Current trends in theory and research* (Vol. 2, pp. 381-403 ). New York: Academic Press.

Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: Reactions to test. *Journal of Personality and Social Psychology, 46*, 929-938.

Sarason, I. G. (1988). Anxiety, self-preoccupation and attention. *Anxiety Research: An International Journal, 1*, 3-8.

Sarason, S., & Mandler, G. (1952). Some correlates of test anxiety. *Journal of Abnormal and Social Psychology, 47*, 810-817.

Sarason, S. B., Davidson, K. S., Lighthall, F. F., Waite, R. R., & Ruebush, B. K. (1960). *Anxiety in elementary school children.* New York: Wiley.

Sax, G., & Collet, L. (1968). An empirical comparison of the effects of recall and multiple-choice tests on student achievement. *Journal of Educational Measurement, 2*, 169-173.

Schunk, D. (1991). Goal setting and self-evaluation: A social cognitive perspective on self-regulation. In P. Pintrich & M. Maehr (Eds.), *Advances in motivation and achievement* (Vol. 7, pp. 85-113). Greenwich, CT: JAI Press.

Scriven, M. (1967). *The methodology of evaluation* (Vol. 1). Chicago: Rand McNally.

Slavin, R. E. (1987). Mastery learning reconsidered. *Review of Educational Research, 57*, 175-214.

Sly, L. (1999). Practice tests as formative assessment improve student performance on computer managed learning assessments. *Assessment and Evaluation in Higher Education*, 24(3), 339-343.

Sly, L., & Rennie, L. J. (1999a). Computer managed learning: Its use in formative as well as summative assessment. In M. Danson, & R. Sherrat (Eds.), *Proceedings of Flexible Learning: Third annual computer assisted assessment conference* (pp. 179-189). Loughborough University.

Sly, L., & Rennie, L. J. (1999b). Computer managed learning as an aid to formative assessment in higher education. In S. Brown, J. Bull, & P. Race (Eds.), *Computer Assisted Assessment of Students* (pp. 113-120). London: Kogan Page.

Smith, R. J., Arnkoff, D. B., & Wright, T. L. (1990). Test anxiety and academic competence: A comparison of alternative models. *Journal of Counseling Psychology, 37*(3), 313-321.

Spielberger, C. D. (1966). Theory and research on anxiety. In C. D. Spielberger (Ed.), *Anxiety and behavior* (pp. 3-20). New York: Academic Press.

Spielberger, C. D. (1972). Anxiety as an emotional process. In C. D. Spielberger (Ed.), *Anxiety: Current trends in theory and research* (Vol. 1, pp 484-491). New York: Academic Press.

Spielberger, C. D. (1980). *Preliminary professional manual for the test anxiety inventory.* Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D., Gonzales, H. P., Taylor, C. J., Algaze, B., & Anton, W. D. (1978). Examination stress and test anxiety. In C. D. Spielberger & I. G. Sarason (Eds.), *Stress and anxiety* (Vol.5, pp. 167-191). Washington, DC: Hemisphere/Wiley.

Spielberger, C. D., in collaboration with Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *State-Trait Anxiety Inventory (Form Y).* Palo Alto, CA: Mind Garden, Inc.

Spielberger, C. D., & Vagg, P. R. (1995). Test anxiety: A transactional process model. In C. D. Spielberger & P. R. Vagg (Eds.), *Test anxiety: Theory, assessment, and treatment* (pp. 3-13). Washington, DC: Taylor & Francis.

Stanford, J. D., & Cook, H. P. (1987). Computer managed learning - its applications to increase student achievement using formative self-assessment. In *J.* Barrett, & J. Hedberg (Eds.), *Using Computers Intelligently in Tertiary Education. A collection of papers presented to Australian Society for Computers in Learning* ( pp. 367-378). Sydney, Australia.

Stobart, G., & Gipps, C. (1998). The under achievement debate: Fairness and equity in assessment. *British Journal of Curriculum and Assessment, 8*(3), 39-45.

Svensson, L. (1977). On qualitative differences in learning: Study skill and learning. *British Journal of Educational Psychology, 47*, 233-243.

Thornton, D., Kevill, R. & Sly, L. (2000). Moving the focus - The challenge of moving an organisation towards a more learner-centered flexible learning environment. In C. M. Wang, K. P. Mohanan, D. Pan, & Y. S. Chee (Eds.), *THLE Symposium. First Symposium on Teaching and Learning in Higher Education. Proceedings* (pp. 33-38). Singapore: Centre for Development of Teaching and Learning, National University of Singapore.

Tobias, S. (1985). Test anxiety: Interference, defective skills, and cognitive capacity. *Educational Psychologist, 20*, 135-142.

Torkzadeh, G., & Angulo, I. E. (1992). The concepts and correlates of computer anxiety. *Behaviour and Information Technology, 11*(2), 99-108.

Traub, R. E. (1992). On the equivalence of traits assessed by multiple choice and constructed response test. In R. Bennet & W. Ward (Eds.), *Construction versus choice in cognitive measurement* (pp. 29-44). Hillsdale, N.J: Lawrence Erlbaum Associates.

Trollip, S. R. (1992). Testing: The forgotten component of instruction. In B. Chia, R. Pennell, & R. Sims (Eds.), *Proceedings of the 1992 Conference of the Australian Society for Computers in Learning in Tertiary Education: A future*

*promised* (pp. 6-10). Sydney, Australia: Australian Society for Computers in Learning in Tertiary Education.

Warner, H., & Thissen, D. (1993). Combining multiple-choice and constructed-response test scores; Towards a Marxist theory of test construction. *Applied Measurement in Education, 6*, 103-118.

Webster, F. (1992). Thinking strategically: a subject group's attempts to maintain quality with reduced resources. In G. Gibbs & A. Jenkins (Eds.), *Teaching Large Classes in Higher Education: How to Maintain Quality with reduced Resources* (pp. 138-147 ). London: Kogan Page Limited.

Weil, M., & Rosen, L. (1995). The psychological impact of technology from a global perspective: A study of technological sophistication and technophobia in university students from 23 countries. *Computer in Human Behaviour, 11*(1), 95-133.

Williams, J. E. (1996). Gender related worry and emotionality: Test anxiety for high-achieving students. *Psychology in the Schools, 33*(2), 159-62.

Wine, J. (1971). Test anxiety and the direction of attention. *Psychological Bulletin, 76*, 92-104.

Worthington, V. L., & Zhao, Y. (1999). Existential computer anxiety and changes in computer technology: What past research on computer anxiety has missed. *Journal of Educational Computing Research, 20*(4), 299-315.

Zakrzewski, S., & Bull, J. (1998). The mass implementation and evaluation of computer-based assessments. *Assessment and Evaluation in Higher Education, 23*(2), 141-152.

Zimmerman, B. J. (1989a). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology, 81*, 329-339.

Zimmerman, B. (1989b). Models of self-regulated learning and academic achievement. In B. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement. Theory, research and practice* (pp. 1-25). New York: Springer-Verlag.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25*(1), 3-17.

Zimmerman, B. J. (1994). Dimensions of academic self-regulation: A conceptual framework for education. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance* (pp. 3-21). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology, 82*, 51-59.

Zoller, U., & Ben-Chaim, D. (1988). Interaction between examination type, anxiety state, and academic achievement in college science: An action-orientated research. *Journal of Research in Science Teaching, 26*(1), 65-77.

# APPENDIX A

# EXAMPLE OF A TESTBANK ANALYSIS REPORT

+Report: 5.5.1.1 Analysis.Questions.Multiple_Choice.Detailed
Page 1

for Testbank PSY114

29-NOV-1999

| MMOQQ-A | LL | #ISSUED | 1 | 2 | 3 | 4 | 5 | 6 | %CORRECT | < 25 | 25-75 | >75 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8101 F | 1 | 72 | 0 | 1 | 66* | 5 | 0 | 0 | 91 | | | * |
| 8102 F | 64 | 33 | 5* | 15 | 5 | 8 | 0 | 0 | 15 | * | | |
| 8103 F | 1 | 57 | 0 | 2 | 51* | 4 | 0 | 0 | 89 | | | * |
| 8104 F | 1 | 72 | 1 | 2 | 69* | 0 | 0 | 0 | 95 | | | * |
| 8105 F | 64 | 22 | 0 | 21* | 0 | 1 | 0 | 0 | 95 | | | * |
| 8106 F | 64 | 28 | 13* | 7 | 3 | 5 | 0 | 0 | 46 | | * | |
| 8107 F | 64 | 19 | 0 | 18* | 0 | 1 | 0 | 0 | 94 | | | * |
| 8108 F | 64 | 23 | 0 | 0 | 23* | 0 | 0 | 0 | 100 | | | * |
| 8109 F | 1 | 76 | 17 | 0 | 1 | 58* | 0 | 0 | 76 | | | * |
| 8110 F | 64 | 18 | 5 | 5* | 1 | 7 | 0 | 0 | 27 | | * | |
| 8111 F | 64 | 23 | 7 | 7 | 6* | 3 | 0 | 0 | 26 | | * | |
| 8112 F | 1 | 91 | 19 | 43* | 6 | 23 | 0 | 0 | 47 | | * | |
| 8113 F | 1 | 75 | 16 | 56* | 3 | 0 | 0 | 0 | 74 | | * | |
| 8114 F | 1 | 69 | 64* | 4 | 1 | 0 | 0 | 0 | 92 | | | * |
| 8115 F | 64 | 23 | 3 | 4 | 3 | 13* | 0 | 0 | 56 | | * | |

MMOQQ – represents the item, by module (MM), objective (O) and number (QQ), that is, 8102, is the second item contained in the first objective of module 8.

F – designates the first attempt by any student.

LL – represents the lock level from which questions are drawn; the type of security associated with a question.

The numbered headings 1 to 6 represent the multiple-choice answer and distractors for each question. The number of times each response has was selected is given and the correct answer is indicated by *.

# APPENDIX B

# SAMPLE CML TEST (FIRST PAGE ONLY)

```
+ 30 Question Exam                    Student.PROCEED
Page   1
                        for Course PSY114       Subject PSY114
21-OCT-1999

       ID: 999999      Time: 10:17  Ent: E02     Try:  1  Seed:
362171004

 *** This is a timed exam and must be returned by 21-OCT-1999 11:17

        Multiple-choice
        For the following questions please answer (1-6) or (a-f)

 ( 1)   1. Question =    8104
           If you need to produce an item from memory, you are
           engaging in a(n) __________ task.
           a) paired associate
           b) recognition
           c) recall
           d) episodic

 ( 1)   2. Question =    8101
           As a process, __________ refers to the dynamic
           mechanisms associated with the retention and retrieval of
           information about past experiences.
           a) primacy
           b) conceptualizing
           c) memory
           d) recall

 ( 1)   3. Question =    8119
           Memories that are unavailable seconds after exposure to
           the information are not likely to have survived
__________
           memory to enter the short-term store.
           a) transient
           b) sensory
           c) decayed
           d) flash-bulb

 ( 1)   4. Question =    8171
           Saul Sternberg used reaction times to infer the nature of
           retrieval from __________ memory.
           a) long-term
           b) short-term
           c) sensory
           d) implicit
```

# APPENDIX C

## SAMPLE ERROR SUMMARY

```
+Report 4.3.1                Student.Proceed.Diagnostic
Page   1
                          for Course PSY114        Subject PSY114
21-OCT-1999

      ID: 999999        Time: 10:39  Ent: E02      Try:  1  Seed:
362171004

                    Percent    S->Student Response
 Ques.   Mod-Obj     Lost      C->Correct Answer

   8.       8-2      3.33      S->c
                               C->d
  10.       8-2      3.33      S->b
                               C->c
  13.       9-1      3.33      S->b
                               C->c
  19.       9-2      3.33      S->c
                               C->b
  20.       9-2      3.33      S->a
                               C->c
  23.      10-1      3.33      S->a
                               C->d
  29.      10-2      3.33      S->c
                               C->b
  30.      10-2      3.33      S->a
                               C->d

     Your score on test #  1 on E02   is  73.33
```

The error summary lists those items that a student answered incorrectly. The "Mod-Obj" column links the item with the module and objective it was selected from. The "Percent Lost" column represents the percentage the student lost for answering that particular item incorrectly.

# APPENDIX D

# ATTITUDE TO CML – STUDY 1

**PLEASE ANSWER ALL QUESTIONS**

(Tick box)

1. Mode of study:
   [ ] Full-time
   [ ] Part-time

2. Previous study of Economics:
   [ ] None
   [ ] TEE or equivalent
   [ ] Economics (Micro) 100
   [ ] Other (please specify)

   ---

3. Sex
   [ ] Male
   [ ] Female

4. Which of the following do you think are the BEST features of computer managed learning (CML) testing on this course? Tick as many or as few options as you wish. If you want, you may provide additional options on the lines provided.

[ ] Immediate feedback
[ ] Freedom to choose a convenient time for your test
[ ] Encourages thorough and continuous revision
[ ] Highlights weaknesses for remedial action
[ ] "Resit" opportunity if you fail a test at the 1st attempt
[ ] Removes the concentrated pressure of a single mid-semester test

You may add additional good features, if you wish:

_____

_____

5. Do you agree with a resit for a failed test
   [ ] Yes
   [ ] No

## APPENDIX D (continued)

6. Do you think that resits should be offered to ALL students

[ ] Yes
[ ] No

7. The practice test before the first test was

[ ] Not helpful
[ ] Helpful
[ ] Very helpful

8. Would you like an optional practice test before each test

[ ] Yes
[ ] No

9. The following are some suggestions for different forms of assessment for 20% of the course mark. Indicate the strength of your preference by giving the one you like most number 1, followed by number 2 for your second preference, if you have one. You do not need to put a number in each box.

[ ] CML Tests worth 20%
[ ] A mid-semester test worth 20% to replace the CML tests
[ ] A second essay worth 20% to replace the CML tests
[ ] No preference
[ ] Other: _____

10. What suggestions can you make for improvements to the CML testing ?

_____

_____

_____

_____

Thank you for your co-operation
For educational use only

# APPENDIX E

## STUDENT DEMOGRAPHICS – STUDY 2

The purpose of this survey is to analyse results for the purpose of improving the effectiveness of CML

**PLEASE ANSWER ALL QUESTIONS**

**Student Number:** .....

| | (Tick box) | | |
|---|---|---|---|
| 1. Mode of study: | 1 | [ ] | Full-time |
| | 2 | [ ] | Part-time |
| 2. Previous study of Economics: | 1 | [ ] | None |
| | 2 | [ ] | TEE or equivalent |
| | 3 | [ ] | Economics (Micro) 100 |
| | 4 | [ ] | Other (please specify) |
| | | | --- |
| 3. Sex | 1 | [ ] | Male |
| | 2 | [ ] | Female |
| 4. Did you sit the CML practice test (P01) | 1 | [ ] | Yes |
| | 2 | [ ] | No |

5. What percent do you **EXPECT** to get on the S02 test that you are about to sit (use a whole number) [ ]

| | | | |
|---|---|---|---|
| 6. My TEE score was | 1 | [ ] | I did NOT get a TEE score |
| | 2 | [ ] | under 300 |
| | 3 | [ ] | 300 to 349.9 |
| | 4 | [ ] | 350 and greater |

Answer question 7 only if you did TEE Economics

7. My scaled TEE Economics mark was [ ]
(use a whole number)

Thank you for your co-operation
For educational use only

## APPENDIX F

## ANXIETY QUESTIONNAIRE – STUDIES 3 AND 5

# APPENDIX G

## CML USAGE – STUDY 5

The purpose of this questionnaire is to gain information that can be used to improve computer managed learning. It is entirely optional but your help would be greatly appreciated.

For the following questions please circle the number that represents your opinion.

| | | | | | |
|---|---|---|---|---|---|
| | Hate it | | | | Like it a lot |
| Q.1 How much do you like computer managed learning (CML) as a form of assessments? | 1 | 2 | 3 | 4 | 5 |
| | Not very well | | | | Very well |
| Q.2 How well do you think CML assessment measures your own understanding of this unit? | 1 | 2 | 3 | 4 | 5 |
| | Not at all difficult | | | | Very difficult |
| Q.3 How difficult is CML to use? | 1 | 2 | 3 | 4 | 5 |
| | No | | | | Yes |
| Q.4 Did you sit the practice test? | 1 | | | | 2 |

If you answered **NO** go to Q.7

| | | | | | |
|---|---|---|---|---|---|
| | Not at all | | | | A great deal |
| Q.5 How much did you study for the practice test? | 1 | 2 | 3 | 4 | 5 |
| | Not at all | | | | Very helpful |
| Q.6 Do you think that doing the practice test was helpful to you in Why or why not | 1 | 2 | 3 | 4 | 5 |

---

# APPENDIX G (cont)

**Omit Q.7 if you answered Yes to Q.4**

Q.7 If you did not sit the practice test was it because you

| | |
|---|---|
| Didn't have enough time | 1 |
| Forgot | 2 |
| Didn't think it would be helpful | 3 |
| other | 4 |

---

| | No | Yes |
|---|---|---|
| Q.8 Have you used the CML system for any subject before this second semester Accounting subject? | 1 | 2 |

Please write any other comments about your CML assessment in this unit.

---

# APPENDIX H

## CML USAGE – STUDY 6

For the following questions please circle the number that represents your opinion.

| | | | | | |
|---|---|---|---|---|---|
| | Hate it | | | | Like it a lot |
| Q.1 How much do you like computer managed learning (CML) as a form of assessments? | 1 | 2 | 3 | 4 | 5 |
| | Not very well | | | | Very well |
| Q.2 How well do you think CML assessment measures your own understanding of this unit? | 1 | 2 | 3 | 4 | 5 |
| | Not at all difficult | | | | Very difficult |
| Q.3 How difficult is CML to use? | 1 | 2 | 3 | 4 | 5 |
| | No | | | | Yes |
| Q.4 Did you sit the practice test? | 1 | | | | 2 |

If you answered **NO** go to Q.7

| | | | | | |
|---|---|---|---|---|---|
| | Not at all | | | | A great deal |
| Q.5 How much did you study for the practice test? | 1 | 2 | 3 | 4 | 5 |
| | Not at all | | | | Very helpful |
| Q.6 Do you think that doing the practice test was helpful to you in Why or why not | 1 | 2 | 3 | 4 | 5 |

_____

# APPENDIX H (cont)

**Omit Q.7 if you answered Yes to Q.4**

Q.7 If you did not sit the practice test was it because you

| | |
|---|---|
| Didn't have enough time | 1 |
| Forgot | 2 |
| Didn't think it would be helpful | 3 |
| other | 4 |

---

| | No | Yes |
|---|---|---|
| Q.8 Have you used the CML system for any subject before this second semester Accounting subject? | 1 | 2 |

Please write any other comments about your CML assessment in this unit.

---

# APPENDIX I

## FEEDBACK SURVEY – STUDY 9

*The purpose of this questionnaire is to gain information that can be used to improve computer managed learning. It is entirely optional but your help would be greatly appreciated.*

For the following questions, if a number is given please circle the number that represents your opinion, otherwise circle the appropriate word.

| | | | | | |
|---|---|---|---|---|---|
| **Q.1** Did you sit the first practice test (P1)? | **No** | | | | **Yes** |
| | **Not at all** | | | | **A great deal** |
| **Q. 2** How much did this first practice test help you to prepare for the first assessed test (E1)? | 1 | 2 | 3 | 4 | 5 |

How or how not?

---

| | | | | | |
|---|---|---|---|---|---|
| | **Not at all** | | | | **A great deal** |
| **Q. 3** How much study did you do for the first practice test? | 1 | 2 | 3 | 4 | 5 |
| **Q.4** You were given the correct answer for any items you got wrong on the first practice test (P1). Did this help you? | 1 | 2 | 3 | 4 | 5 |

If so, how did it help?

---

| | | | | | |
|---|---|---|---|---|---|
| | **None at all** | | | | **A great deal** |
| **Q.5** How much revision did you do between the first practice test and the first assessed test (E1)? | 1 | 2 | 3 | 4 | 5 |
| **Q. 6** Did you sit the second practice test (P2)? | **No** | | | | **Yes** |

***If you answered "No" to Q 6 go to Q13***

| | **Not at all** | | | | **A great deal** |
|---|---|---|---|---|---|
| **Q.7** How much did this second practice help you to prepare for the second assessed test (E2)? | 1 | 2 | 3 | 4 | 5 |
| How or how not? | | | | | |
| | **None at all** | | | | **A great deal** |
| **Q.8** How much did you study for the second practice test? | 1 | 2 | 3 | 4 | 5 |
| **Q.9** How much revision did you do between the second practice test and the second assessed test (E2)? | 1 | 2 | 3 | 4 | 5 |

***Skip Q10 &11 if you did not sit BOTH practice tests***

| | | |
|---|---|---|
| **Q.10** Did you study more after one practice test than the other? | **No** | **Yes** |
| **If you answered "Yes" to Q10** | | |
| Which practice test did you study more after? | **P1** | **P2** |
| What was the reason for this? | | |

---

| | | |
|---|---|---|
| **Q.11** On the first practice test (P1) you were given the correct answer to any item you answered incorrectly. On the second practice test (P2) you were just told which items you got wrong. | **P1** | **P2** |
| Which method was more useful? | | |
| Please explain why | | |

---

| | | |
|---|---|---|
| **Q.12** After the second practice test did you look up the correct answers to items you answered incorrectly? | **No** | **Yes** |

**Q.13** Please write any comment you have regarding the information about incorrect answers and how you use or don't use this information.

_____

_____

**Q.14** Please write any other comment you think may be useful for improving the CML testing system.

_____

_____

**Q.15** My sex is **Male** **Female**

**STUDENT ID:** _____

# APPENDIX J

## FEEDBACK SURVEY – STUDY 10

*The purpose of this questionnaire is to gain information that can be used to improve computer managed learning. It is entirely optional but your help would be greatly appreciated.*

For the following questions, if a number is given please circle the number that represents your opinion, otherwise circle the appropriate word.

| | | | | | |
|---|---|---|---|---|---|
| **Q.1** Did you sit the first practice test (P1)? | **No** | | | | **Yes** |
| | **Not at all** | | | | **A great deal** |
| **Q. 2** How much did this first practice test help you to prepare for the first assessed test (E1)? How or how not? | 1 | 2 | 3 | 4 | 5 |

---

| | | | | | |
|---|---|---|---|---|---|
| | **Not at all** | | | | **A great deal** |
| **Q. 3** How much study did you do for the first practice test? | 1 | 2 | 3 | 4 | 5 |
| **Q.4** You were given the correct answer for any items you got wrong on the first practice test (P1). Did this help you? If so, how did it help? | 1 | 2 | 3 | 4 | 5 |

---

| | | | | | |
|---|---|---|---|---|---|
| | **None at all** | | | | **A great deal** |
| **Q.5** How much revision did you do between the first practice test and the first assessed test (E1)? | 1 | 2 | 3 | 4 | 5 |
| **Q. 6** Did you sit the second practice test (P2)? | **No** | | | | **Yes** |

***If you answered "No" to Q 6 go to Q13***

| | Not at all | | | | A great deal |
|---|---|---|---|---|---|
| **Q.7** How much did this second practice help you to prepare for the second assessed test (E2)? | 1 | 2 | 3 | 4 | 5 |

How or how not?

---

| | None at all | | | | A great deal |
|---|---|---|---|---|---|
| **Q.8** How much did you study for the second practice test? | 1 | 2 | 3 | 4 | 5 |
| **Q.9** How much revision did you do between the second practice test and the second assessed test (E2)? | 1 | 2 | 3 | 4 | 5 |

***Skip Q10 &11 if you did not sit BOTH practice tests***

| | | |
|---|---|---|
| **Q.10** Did you study more after one practice test than the other? | **No** | **Yes** |
| **If you answered "Yes" to Q10** | | |
| Which practice test did you study more after? | **P1** | **P2** |

What was the reason for this?

---

| | | |
|---|---|---|
| **Q.11** After the second practice test did you look up the correct answers to items you answered incorrectly? | **No** | **Yes** |

**Q.12** Please write any comment you have regarding the information about incorrect answers and how you use or don't use this information.

---

**Q.13** Please write any other comment you think may be useful for improving the CML testing system.

---

| | | |
|---|---|---|
| **Q.14** My sex is | **Male** | **Female** |

**STUDENT ID:** _____

# APPENDIX K

## FEEDBACK SURVEY – STUDY 11

Dear Student

The purpose of this questionnaire is to analyze areas where the CML Lab can improve the testing process. To assist in this process would you please answer the following questions and wherever possible add any additional comments you wish. When you had a first attempt at this CML test you received an "error summary" that told you which items you answered incorrectly. It gave you the correct answer to any incorrect response.

| | | | |
|---|---|---|---|
| 1 | On your **first** attempt at this test did you go through the items you answered incorrectly before you left the CML Lab? | **YES** | **NO** |
| 2 | Was it useful to know the correct answer for a item you got wrong? | **YES** | **NO** |
| 3 | If you were only told that your response was incorrect would you try to find the correct answer for yourself? | **YES** | **NO** |
| 4 | If there was a textbook in the CML Lab for you to check correct answers for any item you got wrong would you use it before you left the CML Lab? | **YES** | **NO** |
| 5 | Do you think it would be better ONLY to know that your answer was wrong so that you had to FIND the correct answer for yourself? | **YES** | **NO** |
| 6 | Did you revise before you attempted the test a second time? | **YES** | **NO** |

7 How do you use the information about incorrect answers that you get from the CML system to help your preparation for the next CML test?

---

Any comments you can make regarding the usefulness, or otherwise, of the information the CML system gives you when it marks your answers would be helpful

---

Thank you for your help.

# APPENDIX L

## FEEDBACK SURVEY – STUDY 12

Dear Student

The purpose of this questionnaire is to analyze areas where the CML Lab can improve the testing process. To assist in this process would you please answer the following questions and wherever possible add any additional comments you wish.

1 Did you do the first practice test (P1)? **YES** **NO**

**(If you answered "NO" go to question 4)**

When you did the CML practice test you received an "error summary" that told you which items you answered incorrectly. It gave you the correct answer to any incorrect response.

2 After the practice test did you go through the items you answered incorrectly before you left the CML Lab? **YES** **NO**

3 Was it useful to know the correct answer for a item you got wrong? **YES** **NO**

4 If you were only told that your response was **incorrect** would you try to find the correct answer for yourself? **YES** **NO**

5 If there was a textbook in the CML Lab for you to check correct answers for any item you got wrong would you use it before you left the CML Lab? **YES** **NO**

6 Do you think it would be better ONLY to know that your answer was wrong so that you had to FIND the correct answer for yourself? **YES** **NO**

7 Did you revise before you attempted the "real" test? **YES** **NO**

8 How do you use the information about incorrect answers that you get from the CML system to help your preparation for the next CML test?

---

Any comments you can make regarding the usefulness, or otherwise, of the information the CML system gives you when it marks your answers would be helpful

---

Thank you for your help

# APPENDIX M

## FEEDBACK SURVEY – STUDY 13

Dear Student

This questionnaire aims to find areas where the CML Lab can improve the testing process. Please answer the following questions and add additional comments where you want. You may write on the back of the page, if you wish.

| | | | |
|---|---|---|---|
| 1 | Did you do the first practice test (P1)?<br>**(If you answered "NO" go to question 4)** | **YES** | **NO** |
| 2 | After the practice test did you go through the items you answered incorrectly before you left the CML Lab? | **YES** | **NO** |
| 3 | Was it useful to know the correct answer for a item you got wrong? | **YES** | **NO** |
| 4 | If you were only told that your response was **incorrect** would you try to find the correct answer for yourself? | **YES** | **NO** |
| 5 | If there was a textbook in the CML Lab for you to check correct answers for any item you got wrong would you use it before you left the CML Lab? | **YES** | **NO** |
| 6 | Do you think it would be better ONLY to know that your answer was wrong so that you had to FIND the correct answer for yourself? | **YES** | **NO** |
| 7 | Did you revise before you attempted the "real" test? | **YES** | **NO** |

8 How do you use the information about incorrect answers that you get from the CML system to help your preparation for the next CML test?

---

9 If the "real" (E1) has 30 items and you are able to have some say in the design of the practice test, how many items would it have? (Please circle)

**10 15 20 25 30 35 40**

Please give reasons for your choice

---

Any comments you can make regarding the usefulness, or otherwise, of the information the CML system gives you when it marks your answers would be helpful

_____

Please write any other comments about your CML assessment in this unit.

_____

Thank you for your help. STUDENT ID: _____