# Automating Assessment and Marking of Spreadsheet Concepts

**Peter Summons[1], Jo Coldwell[2], Frans Henskens[3], Christine Bruff[4]**
**Information Systems Group**
**Department of Management**
**University of Newcastle**

**This paper discusses the use of methods for automating the testing and marking of spreadsheet concepts within an introductory service subject offered at the University of Newcastle. The methods reduce logistic and management problems encountered with large classes and enhance individual student learning outcomes. Future extensions to the present work are also described.**

**Key Words: Mastery level, Automated Assessment**

## Introduction

An academic goal of a university is to ensure that student learning outcomes are achieved to their maximum potential. Attainment of academic goals however, can often be in conflict with attempts to realise administrative or business goals, as these may be perceived as being easily achieved through rationalisation of academic costs. The modern university must find a balance in the optimisation of these goals. This was recognised by (Cranitch 1991) "...with large numbers of students with different learning styles, there is a need to develop instructional arrangements that maximise student learning, while trying to minimise the cost in terms of time, effort and money". Practically this may mean trying to maximise individual learning outcomes within very large classes, while trying to cope with the problems associated with large classes. One method of doing this is to reduce the logistic load or the management complexity of large classes, eg the PACE system (Oliver & Mitchell, 1996) for assessment management. Another is by automating some of the tasks related to student learning outcomes, eg the provision of *consistent*, *adequate* and *timely* feedback.

We have developed methods to improve the management and alleviate the logistical burden of assessment of large classes. The methods also maximise individual student learning outcomes. Management of large classes is improved by providing standardisation of marking, by early detection of problems in assessment material, by achieving a reduction in marking time (and quick assessment turnaround), by minimising cheating or unfair student advantage, as well as by automating the compilation, storage and reporting of assessment results. Students' learning potential is increased by providing feedback from the assessment in the form of a complete break-down of assessment marks, by early detection and correction of any factors that

---

1mgpfs@cc.newcastle.edu.au
2jojo@deakin.edu.au
3mgfah@alinga.newcastle.edu.au
4cgcmb@alinga.newcastle.edu.au

may disadvantage the student, by explicitly specifying student learning expectations and by using the assessment as a learning tool.

We have used the methods for the management of a practical assessment and automated evaluation of spreadsheet concepts in an introductory service subject, INFO101, attracting student numbers of over 1000 per year (Summons et al, 1997).

## Background

In 1995, the assessment of spreadsheet and wordprocessing skills in INFO101 was trialed at the University of Newcastle using the general philosophy of a *mastery level*. Mastery level assessment requires students to attain a specified skill level that indicates an acceptable state of competency in the subject. Assessments were skill-based and utilised hands-on, practical tests conducted during students normal tutorials. They were also intended to serve as learning tools. A sample assessment enabled students to practice tasks before their actual assessment. It both demonstrated course expectations and guided student learning in line with Ablong's (1991) view that "the best way to learn is by doing".

In 1996, the trial was extended to include automation of the marking for the INFO101 Excel assessment. The aims of the 1996 trial were to:
- standardise the assessment criteria;
- standardise the assessment evaluation;
- rationalise the resource allocation in assessment preparation and marking.

## Assessment Rationale

A primary goal of mastery testing was to establish a measure for the presence of a predetermined competency level in students. Underlying assumptions were that the specified competency level could be demonstrated by an individual's attainment of particular skills, and that their mastery of those skills could be ascertained by means of their performance in predetermined tasks. Implicit in the latter assumption was that the successful completion of a task was, in fact, associated with the presence of a particular skill. This meant that not only did the completion of the task require the presence of the skill, but also that the degree of completion of the task accurately reflected the level of attainment of the associated skill. The creation of the mastery assessment for INFO101 was therefore perceived as a threefold process:

1. Identify basic skills necessary to attain the overall competency level.
2. Set tasks which demonstrate the appropriate skills.
3. Determine minimum, and higher, levels of successful attainment for the tasks that constitute agreed-upon measures of competency in a particular task, and hence in the skill that was demonstrated by its achievement.

In effect, a taxonomy of educational objectives was identified, similar to that established by Bloom (in Stones 1979). We focussed on Bloom's *Knowledge Level* and *Comprehension Level* in his Cognitive Domain taxonomy to identify generic

spreadsheet concepts for assessment.  These were considered to be core skills and were used to set basic competency levels.  Examples included the ability to correctly:

- navigate around the worksheet;
- format data;
- use COPY versus MOVE operations;
- use ABSOLUTE versus RELATIVE addressing;
- create FUNCTIONS, eg SUM, AVERAGE.

These basic skills corresponded to *Knowledge-Level* cognitive objectives, requiring memory involving *simple-knowledge* and *knowledge-of-a-process* (Cangelosi 1992).

Additional skills, considered important in the degrees for which INFO101 was a compulsory subject, were included in the 1996 assessment.  These were more oriented towards Cangelosi's *Intellectual-Level* cognition, requiring the ability to reason and make judgments.  They included the ability to:

- model simple financial function applications eg PMT;
- use the IF function for simple applications;
- use combinations of functions.

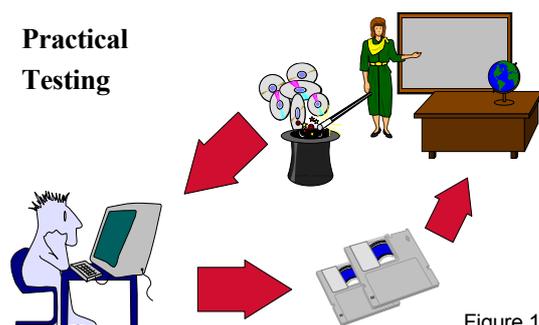## Assessment Methods and Management

**Practical Testing**



Figure 1

Computer Based assessment packages for Excel were designed to enable both easy testing and to eliminate the collusion commonly found in assignment-type assessment of the large student numbers involved in INFO101.    Figure 1 demonstrates the assessment process. Many versions of the practical assessment tasks are placed on our Novell computer network and "hidden" from general access.  A tutor "unhides" one version and the tutorial class undertakes an open-book, hands-on assessment with it.  The students give their disks (assessment answers) to the tutor, who marks their assessments and returns the marked disks to the students.

A computer package was designed to automate the marking of the 1996 INFO101 Excel assessment.    The marking package uses Excel macros (Visual Basic procedures) to assess individual student answers to the assessment.  Figure 2 illustrates the marking process for one student in one tutorial group.  The marking program (Marker.xls) prompts for the disk containing a student's assessment answer (Student.xls), automatically evaluates the student's answer and generates appropriate marks for each individual assessment task, as well as an appropriate overall grade for each student.  The grade is displayed to the tutor responsible for marking the assessment, allowing for manual checking if required.
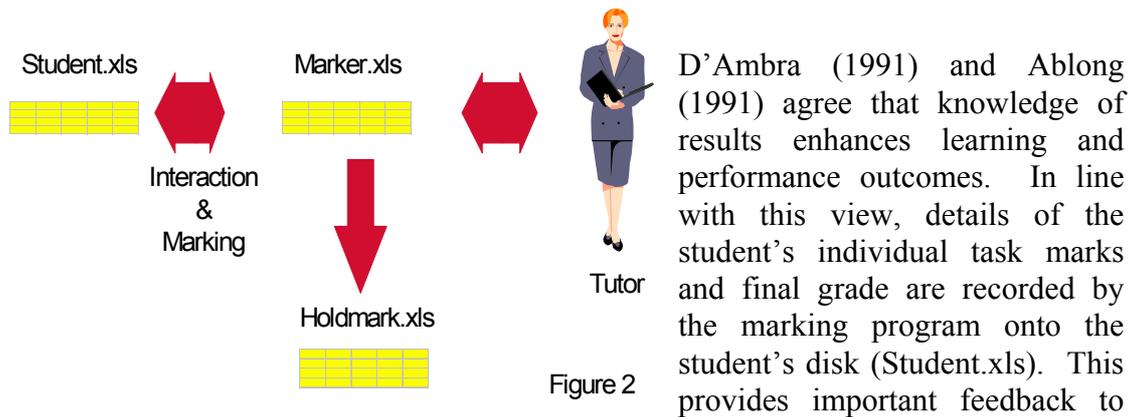
Student.xls    Marker.xls

Interaction
&
Marking

Holdmark.xls

Tutor

Figure 2

D'Ambra (1991) and Ablong (1991) agree that knowledge of results enhances learning and performance outcomes. In line with this view, details of the student's individual task marks and final grade are recorded by the marking program onto the student's disk (Student.xls). This provides important feedback to students, as they can then clearly identify specific problem areas. The marking package also records details for the tutor, appending each student's name and student number, the test version they had attempted, their individual task marks and their final grade onto a marks file (Holdmark.xls) for storage. The "marks file" is actually an Excel worksheet that acts as the mark repository for a tutorial class. This file is later combined with marks files from other tutorials groups, producing a composite mark file for the subject.

The marking package automatically checks each assessment task and awards marks according to a pre-set marking scheme. This can be accomplished due to the layout constraints and specificity of the required answers.

An exam question is considered to be of poor quality if it attempts to test too much, ie more than what is intended by the examiner. An example may be an examiner trying to test whether the concept of formulating the VLOOKUP spreadsheet function correctly had been achieved. If the examiner sets a question that has underlying assumptions regarding the mathematical or financial skills of the student, then the student answer may hide their ability to correctly formulate a VLOOKUP function and confuse it with their financial or mathematical ability. Our assessment strategy requires the examiner to focus on what they intend to test, and how to mark (test) it, well before they set the question. We attempt to ensure that our assessment questions actually test what is intended.

To illustrate, one task required the student to:
"*Give Row 3 a Background colour of YELLOW with a BLACK cell border line underneath it*"
The marking code for this task selected row 3 and did the following test:

```
If Selection.Interior.ColorIndex <> 36 And Selection.Borders(xlBottom).ColorIndex <> 36
        And Selection.Borders(xlBottom).ColorIndex <> Selection.Interior.ColorIndex
    Then
            mark1f = 1
    Else
            mark1f = 0
    EndIf
```

The marking code produced for the task reflected the examiners' requirement that the student knew how to change the specified row colour, and that they could insert a bottom border. It did not distinguish particular colours, or hues, as this would disadvantage colour-impaired students. It made sure that the colour of row 3 had changed from its original colour (ColorIndex 36), and also made sure that the colour

of the bottom border was different to the new row colour and from its previous colour, ie that a border had been inserted.

| a) Use a FUNCTION in cell B6 to calculate the TOTAL of the cells outlined in DARK BLUE | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | 2 | | | | |
| Answer Is: | | | | 33 | | | | |
| | | | | 44 | | | | 9 | 4 |
| | | | | 55 | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | 55 | | | | | |
| | | | 66 | | 12 | 55 | 98 | |
| | | | 32 | | | | | |
| | | | | | | | | |
| b) Left-Justify the contents of the DATA cells that are outlined in DARK BLUE | | | | | | | | |
| Go to NEXT WORKSHEET | | | | | | | | |
| | | | Figure 3 | | | | | |

The assessment tasks had to be carefully produced to suit the aim of the examiner. In the worksheet of Figure 3, if the aim of the task was simply to test that a total could be produced, a simple comparison to find both a formula and the correct value would suffice. If however, it was also an aim to test that the student could perform multiple selections, then the question would need to be modified, eg a block select of all data cells would produce the required value in cell B6, ie =SUM(C3:H12). In this example, the marking code could check that multiple selections are present in the formula, but this can get very messy in more complicated cases, ie with poorly designed questions. A better alternative is to adjust the question to take the testing aim into account, eg Figures 4 shows a redesigned task that will produce a circular reference error if the data is blocked in one operation.

| a) Use a FUNCTION in cell F8 to calculate the TOTAL of the cells outlined in DARK BLUE | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | 2 | | | | |
| | | | | 33 | | | | |
| | | | | 44 | | | | 9 | 4 |
| | | | | 55 | | | | |
| Circular Reference error | | | | | | | | |
| for =SUM(C3:H12) | | | | | Answer Is: | 0.00 | | |
| | | | | | | | | |
| | | | 55 | | | | | |
| | | | 66 | | 12 | 55 | 98 | |
| | | | 32 | | | | | |
| | | | | | | | | |
| b) Left-Justify the contents of the DATA cells that are outlined in DARK BLUE | | | | | | | | |
| Go to NEXT WORKSHEET | | | | | | | | |
| | | | Figure 4 | | | | | |

Assessment tasks can include graduated marking, such as allocating some marks if a portion of a task is completed. This is explored in more detail in Summons et al (1997).

Discriminator questions were also part of the 1996 assessment. They required thought in applying basic skills, and the ability to combine several basic skills for their completion. Individual tasks were not explicitly stated. It was up to the student to decide what appropriate tasks were required. An example is given in Figure 5. If such questions are not included in the sample assessment, they become discovery questions, allowing the examiner to test application skills, ie test an end result rather than the steps that were required to achieve it. The skills involved in the intermediate steps would have been tested previously in other, more specific, questions. In such a case, it would be appropriate to simply compare student answers against expected answers.

| Customer | Trans Type | Part Number | Unit Price | Units Bought | Total Sale Price | Amount of Discount | Discounted Total | Sales Tax Applied | Total Amount |
|---|---|---|---|---|---|---|---|---|---|
| Bill | T | BU120 | $ 2 | 30 | | | | | |
| Sue | R | FF76 | $ 4 | 12 | | | | | |
| Bob | T | HT7899 | $ 5 | 17 | | | | | |
| Kim | T | YU784 | $ 4 | 21 | | | | | |
| Li | R | BU120 | $ 3 | 3 | | | | | |
| | | | | | | | | | |
| Sales Tax | | | 15% | | | Number of Customers: | | | |
| Discount Threshold | | | 20 units | | | Avg Amt of Discount: | | | |
| Discount Percentage | | | 4% | | | Smallest Unit Price: | | | |
| Go to NEXT WORKSHEET | | | | | Figure 5 | | | | |

Tasks that require a qualitative judgement in marking can be included. These tasks incorporate interaction with the tutor in their marking scheme. The 1996 assessment contained one such interactive task, included primarily to see how this approach would be received. The aim of the task was to have the student produce a specified type of chart eg a pie chart, possessing certain properties eg their name as a title, from a given set of data. The marking package displayed the worksheet containing the student's answer for the tutor, as well as creating and displaying a correct version of the chart next to the student's attempt for easy visual comparison. It then accepted a mark from the tutor for the student's work. It should be noted that the task *could* have been marked without tutor interaction, with all checking of chart properties accomplished entirely by the marking package.

Ablong (1991) noted that "the use of definite instructions in a task is important and assists the learner...". The 1996 assessment addressed this by a specificity requirement in questions. We attempted to make the requirements of the questions very direct, and to avoid any unnecessary preamble. As requirements for answers between particular tasks varied, the directions as to where the answers should appear were included with each question. To further minimise ambiguity a standard was adopted that directed students to constrain their answers to a coloured region on the screen.

**Assessment Results**

A sample exam was made available to the students well before their assessment task (approximately 2 weeks). The early timing ensured that potentially all students had an equal preview of their assessment. This was beneficial in reducing advantages arising from collusion or pre-knowledge, where one tutorial class may have gained an advantage over another by doing the assessment significantly later than the other class.

It was found that the sample assessment encouraged student feedback. This allowed early detection and prevention of potential problems associated with the assessment

instrument, such as colour differentiation problems caused by students' physiology or by hardware abnormalities. Early detection of these problems enabled either the rectification or the minimisation of their effects in the real assessment. Students became familiar with the style of language and questions they could expect in their final assessment and this helped minimise ambiguities in questions and mis-interpretation of requirements.

INFO101 tutors were asked to manually mark some student assessments according to a set marking scheme, and to compare their results, such as their time spent marking and the marks they awarded, with the automated marking program. A survey of tutors involved in the 1996 trial discovered that the majority of tutor results would have varied from the marking scheme. Tutors marks also varied between tutors (discrepancies between tutors for the one task in the 1996 assessment that required a qualitative judgement from the tutors were not included). This confirmed that the automation of marks resulted in a more standardised assessment for students.

The tutor survey indicated considerable savings in time from manually marking to automated marking. It was hard to judge time savings accurately as the assessment contained a task that required interaction between the tutor and the marking program for testing purposes, so that automated marking times had a manual component that would normally not appear. Nevertheless, tutors marking time was reduced by an average of 67%. The time-saving, approximately 100 hours marking for INFO101 in 1996, was reflected in a considerable savings in resources (tutors wages), which becomes a recurrent saving in subsequent years.

A major learning outcome of the 1996 trial confirmed that the best approach to a test is a methodically ordered scheme that firstly decides on what skills or outcomes are required to be tested, then defines a method and marking scheme that will adequately test individual skills, and finally generates the question to test an appropriate skill or outcome. A question composed without regard to how it will mark individual skills may not really test the attainment of these skills.

The 1996 pilot scheme for automated marking of Excel resulted in a University of Newcastle CALT teaching grant for 1997. The grant will fund the production of a package to generate multiple versions of tests for future Excel assessments. The package will also mark the assessments and manage the collection, storage and reporting of student marks. The grant is also being used to establish an extensive test bank of graded questions for Excel assessment. The aims of the 1996 trial have now been extended to enhance individual student learning outcomes and improve assessment management of INFO101 by producing a computer package that will:

- Enable standardisation of marking, eliminating any bias existing in inter-tutorial and/or intra-tutorial marking. Given the large number of markers (tutors) involved in INFO101, this is extremely difficult without the package;
- Allow easy articulation for students with existing knowledge of the assessment topic. An initial assessment grade could be applied. This may be accepted by the student, or bettered if they wish to attend the tutorials and attempt the formal examination. This would be logistically impossible otherwise, given the large number of students and the existing resources;

- Minimise cheating and group advantage by allowing all students equal opportunity to view the assessment (via a pre-assessment sample test) and having enough variations of the actual assessment (of a similar difficulty level) to ensure that a students' assessment result is attributable solely to their individual effort;
- Allow easy administration of post assessments to students who miss their initial assessment without prejudice as to the fore-knowledge of the student, or to the grading of the student in the context of the entire class group;
- Provide an early learning stimulus for students, allow for tutor remediation in identified areas and generate detailed feedback to students in the form of individualised question evaluation;
- Be capable of use as a tutorial aid giving individual feedback for postgraduate courses.

## Conclusions and Future Work

Benefits of the use of the mastery concept in assessment have been discussed in the context of a practical trial carried out at the University of Newcastle in 1996. The methods and results of automated marking techniques for assessing mastery of spreadsheet concepts has been outlined. It was seen that the use of pre-assessment practice tasks proved to be beneficial both to learning and to assessment. Students responded well to the quick turnaround and to the comprehensive marks that the automated marking program placed on their assessments. The use of automated marking techniques, even considering that the pilot program contained some interaction with the tutor, were also found to result in significant savings of time and money. Bias, due to inter-tutorial or intra-tutorial marking variations was demonstrated in a survey of manual marking. The elimination of this bias and the standardisation of student marks was proposed through automated marking techniques. Assessment management advantages were also found in the aggregation and storage of course marks.

The current work for 1997 and the anticipated benefits to the assessment of Excel in INFO101 were outlined in the paper. The automated marking concept is being extended to cover the assessment and marking of Access in INFO101, with a trial scheduled for 1997. Investigation into extensions of the concept to the other areas covered by INFO101, eg Word, will also be undertaken. A liaison with Deakin University is currently investigating other areas in which the concept can be applied. The liaison is also investigating the production of a package that can be used in distance learning mode as a tutorial and also as an assessment.

## Bibliography

Ablong, A.P.  Theories of Learning and Their Application to Prototyping *Proceedings of the International Working Conference on Shaping Organisations Shaping Technology*, pp1-11, Adelaide, South Australia, Oct, 1991.

Cangelosi, J.S.  *Systematic Teaching Strategies* Longman Publishing Group, London, 1992.

Cranitch, C.J.  Computer-Based Education: Is it an Effective Alternative to Classroom Education *Proceedings of the Australian Computer Conference ACC'91 MOSAIC*, pp111-122, Adelaide, South Australia, Oct, 1991.

D'Ambra, J.  Computer Literacy - No Longer an Option *Proceedings of the Australian Computer Conference ACC'91 MOSAIC*, pp111-122, Adelaide, South Australia, Oct, 1991.

Oliver, B.K., Mitchell, G.C.  Setting the PASE - The Value of Computer Aided Assessment *Proceedings of the 1st Australasian Conference on Computer Science Education*, University of Sydney, 1996.

Stones, E.  *Psychology of Education: A pedagogical approach* Methuen & Co, London, 1979.

Summons, P.F., Coldwell, J., Henskens, F., Bruff, C.  Automated Evaluation of Spreadsheet Concepts - Excel in Action *Proceedings of the 2nd Annual NSW Symposium on Information Technology and Information Systems SITIS'97*, University of NSW, Sydney, Feb, 1997