IMPROVING EXAMINATIONS IN A UNIVERSITY SCIENCE COURSE: A COOPERATIVE EXERCISE BETWEEN TWO ACADEMICS

Diane J. Grayson[1] and Matthew C. Clarke

University of Natal, Private Bag X01, Scottsville 3209, South Africa

ABSTRACT

Examinations are a significant feature of university life. To a large extent, the whole of a student's life is geared towards passing examinations. For this reason, what students choose to learn will be driven largely by what they expect is going to be in the examination. This is particularly true when the bulk of their final mark is derived from the examination, as is the case in most university courses, particularly in the sciences. Thus it is appropriate that a great deal of thought and care should go into the setting of examinations. However, few academics receive any educational training, including on how best to set examinations. Often examinations are set in a rush, with no pedagogical guidelines available to direct their design and construction. Our concern with this state of affairs prompted us to look into how examinations, and written examinations in particular, might be improved. One of us designed a tool to help the examination writer create a balanced examination paper, well-matched to the goals of a course, and the other used the tool in the setting of his examination papers. In this paper we describe the tool, the experience of using it and give suggestions for how it might be used in future. On the basis of our experience, we reflect on how other aspects of assessment in university science courses might be improved.

INTRODUCTION

For most universities in the world, final examinations provide a significant, if not the predominant, means of assessing students. Heywood (1989, p. 4) quotes a number of assumptions that underlie the use of university examinations, some of which are clearly specious. These include: the assumption that 'quality' of academic performance is rateable on a single continuum from first-class honours to failure; the assumption that examinations are impartial; and the assumption that forced regurgitation of knowledge under stress is predictive of future performance. Whilst examinations may not fulfil all the roles that academics think they do, examinations tend to be treated as a valid, or at least a pragmatic, means of conducting summative assessment. Habeshaw et al. (1993, p. 63) suggest some reasons for this, such as easing the marking load when there are large numbers of students (as opposed to doing continuous assessment), ensuring that students do not copy from one another and so being able to judge the unaided performance of each individual student, and fulfilling the requirements of certain professional bodies.

At the University of Natal, Pietermaritzburg, students' marks are obtained largely from their performance in final examinations. Indeed, up until last year university regulations prohibited the year mark (derived from continuous assessment) from counting more than one

[1] New address: MASTEC, PO Box 55873, Pietersburg 0700, South Africa

third of the final mark. Thus good performance in the examinations is essential if students want to pass their courses. Although academics bemoan the fact that students are only interested in learning something if it will be on the examination, this attitude is hardly surprising given the competitiveness of today's world. On the other hand, academics can use the structure of the examinations as an incentive to encourage students to learn the things that we believe are important to learn. This requires that there be a close match between the goals of a course and what is tested in the examination. As Ramsden (1992, p. 189) puts it, "The teacher with a well-developed understanding of assessment will strive to connect his or her goals for learning firmly with the assessment strategies he or she uses." However, such a match does not always occur. One reason is that few academics have any educational training. Another is that academics do not always appreciate the extent to which the examinations drive student learning. A third is that they may not know how to go about creating an examination paper that closely matches their course goals.

In order to try to address some of the problems mentioned above, the first author, who is a science educator, created a tool that can be used by academics to assist them in constructing examinations that match their course goals. This tool was made available to academics in the Science and Agriculture Faculties at the University of Natal through workshops that she ran. The second author, who is a senior lecturer in the Department of Computer Science and Information Systems, used the tool when he set his final examinations in 1996. In the remainder of this paper, we shall describe the tool, the experience of using it, and our resulting reflections on university examinations and assessment in general.

A TOOL FOR WRITING AND ANALYSING EXAMINATION PAPERS

The examination-writing tool (EWT) was designed on the assumption that for an examination paper to be balanced it should reflect the relative weight given to the various objectives of the associated course and the relative importance given to different types of thinking, from memorisation to problem-solving skills.

Testing objectives

The examination-writing tool comprises two tables, which examiners fill in using their draft examination papers, plus a list of question types. Before filling in Table 1, they write down the objectives of their course, labelling them *a*, *b*, and so on, and the relative weight given to each course objective as a percentage, ensuring that the weightings add up to 100%. The examiner then looks at each question on the examination paper and writes down next to it the letter(s) corresponding to each of the objectives (there may be more than one) being tested in that question. Once this has been done for the whole paper, the total number of marks of all the questions testing objective *a* are written in column 2 of Table 1 next to *a*, the number of marks of questions testing objective *b* are written down next to *b*, and so on for each objective. The marks in the second column are added up to obtain TOTAL 1. This total will almost always exceed the total number of marks on the paper, since it is common for individual examination questions to test more than one objective. The third column in Table 1 is then filled in by dividing each figure in column 2 by TOTAL 1 and converting it to a percentage. The figures in column 3 thus represent the percentage of the paper that is devoted to testing each objective. (See Appendix I for a worked example.) These figures can then be compared with the original list of objectives and relative weightings written down by the examiner before completing Table 1. Any major discrepancies can then be addressed and, if desired, corrected by modifying the examination questions.

TABLE 1

Table for determining the relative weight given to each course objective in the examination

Objective	Marks	% (of TOTAL 1)
a		
b		
etc.		
TOTAL 1		

Different types of questions

In addition to identifying which objectives the questions in an examination assess, it is also useful to identify the type of knowledge and depth of understanding needed to answer each question. We have identified ten types of knowledge and understanding that might be assessed by means of written examination questions, listed in Figure 1. We recognise that this list is not exhaustive, but we think it contains the most common types of questions given in university science examinations. However, we encourage academics who use the EWT to invent additional categories if their questions do not fall readily into our categories.

1. Memory of
   (a) factual information (b) procedures or algorithms
2. Application of remembered factual information to
   (a) known context (b) unknown context
3. Application of remembered procedures or algorithms to
   (a) known context (b) unknown context
4. Application of factual information given in the question to
   (a) known context (b) unknown context
5. Application of procedures or algorithms given in the question to
   (a) known context (b) unknown context
6. Conceptual understanding
7. Proficiency at individual thinking/ reasoning skill(s), e.g. use of analogy, classification, comparing and contrasting, translation between different types of representations, development of a logical argument
8. Proficiency at higher order cognitive skills
   e.g. synthesis, integration of knowledge, appreciating relationships between knowledge acquired in different contexts
9. Proficiency with strategies,
   e.g. problem-solving approaches, techniques
10. Understanding of epistemology of the discipline (how knowledge is created)

Figure 1: Types of examination questions

The first five items in Figure 1 can also be represented in the form of a matrix, which separates two dimensions–the knowledge required to answer the question and the context to which the knowledge must be applied. This matrix is shown in Table 2.

TABLE 2

Matrix for categorising questions that test memory and application of information

Application required	Memorised		Given in the question	
	Factual information	Procedures or algorithms	Factual Information	Procedures or algorithms
Not applied	1a	1b	n/a	n/a
Applied to a known context	2a	3a	4a	5a
Applied to a new context	2b	3b	4b	5b

The types of questions contained in Table 2 (types 1 to 5 from Fig. 1) are illustrated in Appendix II using an example from symbolic logic. Questions of types 6 to 10 from Figure 1 are illustrated in Appendix III using examples from physics. The questions in Appendices II and III are designed to illustrate how different types of questions can be written so that each one tests a different type of knowledge, application or thinking. However, we recognise that in practice one question will often fall into more than one category.

Testing using different types of questions

The second table (Table 3) of the EWT is used to determine how balanced the examination paper is in terms of types of questions. To fill in this table examiners go through their papers again, this time writing down the number next to each question that corresponds to the relevant question type (from Figure 1). The marks of all examination questions corresponding to each question type are then added up and written down in column 2 of Table 3 next to the appropriate question type number. The marks in column 2 are summed to give TOTAL 2. Column 3 is then filled in by dividing each number in column 2 by TOTAL 2 and converting it to a percentage. Thus column 3 indicates the relative weight given to each type of question in the examination paper. Note that although examiners could write down the desirable weightings of each question type before completing Table 3, as was done for the objectives, this is not a straightforward task. It practice, it seems better for examiners to fill out Table 3 first, then look at the figures in column 3 and decide, on the basis of their intuition, whether these figures match well with the philosophy of their courses.

TABLE 3

Table for determining the weight given to each type of question in the examination

QUESTION TYPE	MARKS	% (of TOTAL 2)
1		
2		
etc.		
TOTAL 2		

AN EXPERIENCE OF USING THE EXAMINATION-WRITING TOOL

In this section, the second author describes his experience of using the EWT.

Background

The Department of Computer Science and Information Systems offers two computer literacy courses. One, Computer Science 101 (CS), is offered primarily to science students and the other, End User Computing 110 (EUC), is offered primarily to business students. CS is the older and simpler of the two and I have been involved in teaching it for five years. EUC covers all the topics in CS and also includes broader organisational and commercial issues.

The final exam for both CS and EUC was traditionally a two-hour written paper which counted for 70% of the overall assessment. The other 30% came from six practical assignments and two practical tests. The final exam was marked out of 100, and consisted of 50 one-mark multiple-choice questions (MCQs) and 50 marks' worth of questions requiring written answers of between three and twenty lines. As of last year, the MCQs were administered by computer, each student receiving a random selection of 50 questions from a database comprising the past five years' worth of MCQs.

EUC has recently been drastically restructured on the basis of perceived shortcomings with both CS and EUC, including decreasing the emphasis on examinations. It was therefore particularly appropriate to look critically at the course examinations. While setting final exams for the second semester of 1996, I either applied or supervised the application of the EWT to both CS and EUC, as well as to some past CS exams. The remainder of this section describes the results.

Application of the EWT to Multiple Choice Questions

Since EUC is now significantly different from previous courses, I wanted to ensure that the course objectives were appropriately covered by the computerised database of MCQs. The EUC objectives are described in fine detail using a hierarchy of 126 Intended Learning Outcomes. I ignored those Outcomes which had already been assessed by previous practical tests and noted which of the remaining Outcomes were tested by each of the 301 question in the original database. I deleted 99 questions from the database, either because they referred to concepts which are no longer part of the course, they were too similar to each other, or they required the application of procedures or higher order cognitive skills which were not intended to be assessed by this part of the exam. I also made substantial changes to 11 questions and minor changes to numerous others to improve their clarity.

As expected, when the percentages of questions were calculated in accordance with Table 1, some Outcomes were over-assessed and others were not assessed at all. This tabular approach made it easy to see which topics required adjustment. In order to balance the percentages in Table 1, I added a further 52 questions, bringing the total to 254. This process guaranteed that all questions in the database related to the published course objectives and made the numbers of questions which assessed each objective roughly equal.

The ideas behind Table 3 were also applied to the multiple-choice sections of both CS and EUC, though the whole procedure was not strictly followed. I discovered that the MCQs attempted to test a variety of types of knowledge and understanding. For instance, some presented exercises which required a memorised procedure to be applied to an unknown context (Question type 3b) as in Question 1 of Figure 2. Others assessed how well the student could apply their knowledge to real-life situations (Question type 2b), as well as some deeper understanding (Question types 6 and 7), as in Question 2 of Figure 2.

1. The decimal number 19 is represented in binary, using a 6-bit word as:
    - (a) 010011
    - (b) 011001
    - (c) 100110
    - (d) None of the other answers is correct[2]
    - (e) 110010

2. You have been employed by a car manufacturer who wants to send very detailed manuals of a large number of car models to their agents. All the information is already in the computer. What device should they use to output the information?
    - (a) Laser printer
    - (b) Daisy wheel printer
    - (c) Impact printer
    - (d) Microfilm
    - (e) Ink jet printer
    - (f) Plotter

Figure 2: Examples of old multiple-choice questions given in EUC and CS

It became clear to us that the form of a question must be chosen to be appropriate to the type of knowledge or skill being assessed. Although it is possible to assess more than just

[2] Note that the wording of this differs from the standard “None of the above.” This is required because the computer program which administers this test is set to randomise the order of the answers. The purpose of this is to confound any student attempts to cheat by looking at other students’ screens. This is probably a useful feature, but makes it difficult to write questions whose answers refer to each other. For instance, one cannot use multiple-choice answers such as “Both (a) and (c) are correct.”

memorisation with multiple-choice questions, it is often unclear what students' answers imply about their real understanding. In EUC we have therefore decided that the multiple-choice section should be used exclusively for assessing memory rather than application of knowledge. One of the benefits of the procedure which generates Table 3 was that it highlighted this distinction.

Application of the EWT to Questions with Longer Written Answers

In the past, questions with longer written answers were given in order to test a student's understanding at a deeper level than multiple-choice questions. However, when classified according to the types of questions given in Figure 1, it became apparent that most of these questions are still testing only a student's memory, as shown by the example in Figure 3. In this question, each answer was allocated two marks because it would take the student at least two minutes to understand the question and write the answer, but perhaps the same knowledge could more appropriately have been tested using a multiple-choice question worth just one mark.

> Define the following terms in 2 to 3 lines each [2 marks each]:
> (a) Backing up
> (b) Communications protocol
> (c) etc.

Figure 3: Example of an old longer question given in CS that tests memory only

There were, however, some questions which were more appropriate in the CS paper, as illustrated in Figure 4. These questions require some deeper conceptual understanding (Question type 6) and also assess some higher order reasoning skills (Question type 7). For questions of this type the written answer format is therefore appropriate.

> 1. Compare and contrast magnetic tapes and magnetic disks. [5 marks]
> 2. Briefly discuss the advantages offered by a spreadsheet. [5 marks]

Figure 4: More appropriate old questions given in CS.

Changes that resulted from using the EWT

Through applying the EWT, I gradually became more and more dissatisfied with the structure of the CS exam. Without even applying the full procedure to past exams it was clear that little consideration had ever been given to distributing the marks evenly among the course objectives. In earlier EUC exams, the marks were distributed according to the number of lectures given on each topic. This was the best that could be done at the time because there were no explicitly stated course objectives. Using the EWT also made it immediately clear that the questions which assessed no more than memorisation formed a much larger proportion of the examination than we had imagined.

Our dissatisfaction has led to various changes in the two courses, changes to both the final exams and to the overall assessment strategies. In CS, the number of multiple-choice questions has been reduced from 50 to 30 in order to allow more questions to be asked which require the use of higher order reasoning skills. In addition, the value of the final exam has been reduced from 70% to 40% of the final course mark in order to increase the emphasis on the assessment of the practical objectives via lab tests.[3]

In EUC, more radical changes have been implemented. Continuous assessment (assignments, quizzes, a written theory test and two practical tests) now contribute 70% to the final mark and the final two-hour exam contributes only 30%. The EUC exam is now clearly separated into two parts — a one-hour test of facts, definitions and basic computer concepts, and a separate one-hour test in which these concepts are applied to various real social issues and organisational case studies. The first part is administered using the multiple-choice database described earlier. The second part is written on paper under normal exam conditions, however, it is open-book (i.e. students are allowed to refer to the prescribed textbook during the exam).

The questions in the new EUC written exam are typically structured as case studies. Students are required to apply the information and procedures which they have been taught (and which are described in the textbook) to contexts which they may or may not have seen before. Two examples are shown in Figure 5.

[3] University regulations were changed last year to allow continuous assessment to contribute substantially to the final mark. The regulations were worded in such a way that although 50% of the final mark had to externally examined, that 50% did not have to comprise examinations only - other work done during the course could also be scrutinised by the external examiner. These changes to the university regulations were brought about by pressure from a group of people concerned with educational development in the institution. Thus our experience shows that where regulations constrain pedagogically desirable practices, pressure should be applied to change the regulations, rather than accept the inevitability of having to operate within them.

1. Consider a database which contains the following fields —

Part Number	6 characters
Description	50 characters
Location	30 characters
Quantity	A number which takes up 2 bytes
Unit Price	A number which takes up 8 bytes

For each of the following questions, show all your working.

(a) Suppose the decimal number 472 is entered into the field Quantity. What binary number is actually stored by the computer? [3 marks]

(b) How much disk space (in megabytes) would be required to store this database if it contained 4,000 records? [3 marks]

(c) How many minutes would it take to copy this database to another computer over a communications link at 9600 bps? [3 marks]

2. A staff of secretaries changes from manual typewriters to PCs with word processors. The various files which they type are stored on a common hard disk, and the secretaries are instructed to keep backups of stiffy disks mounted on drives in their own terminals. After a week's work, several documents have been typed, which have been saved on the hard disk as follows —

[A rather disorganised set of directories and files is pictured.]

(a) Show diagrammatically how you would reorganise this disk. [6 marks]

(b) Explain the importance of the backup disks. Speculate on what would happen if the hard disk crashed after three months' operations. [3 marks]

Figure 5: New types of examination questions given in EUC.

Notice that to answer these questions students must demonstrate a deeper and more practical understanding of the basic concepts than is required by the sample questions from CS given earlier. These questions also require students to demonstrate more advanced reasoning skills and an ability to apply their knowledge rather than just to describe what they know. We perceive this to be more in line with the intention of the EUC course. The exam evaluation procedure which underlies the construction of Tables 1 and 3 was instrumental in highlighting these distinctions, indicating the ways in which our past exams fell short of the ideal, and pointing the way towards improvement.

OBSERVATIONS ABOUT THE EWT AND RECOMMENDATIONS FOR ITS FUTURE USE

Table 1 is more objective than Table 3

Table 1 allows for a direct comparison of the actual percentages of marks allocated to each course objective to the ideal, i.e. what the lecturer considered to be the relative importance of each objective. Any major discrepancy can then be addressed. However, Table 3 is more suggestive than prescriptive. The procedure for completing Table 3 does not provide any means by which an ideal mixture of question types can be derived. Hence no comparison can be made between the ideal and actual allocation of marks to question types. The examiner is left to make a gut-level evaluation of whether the percentages "are in line with the philosophy of the course." Although it would be possible to make Table 3 less subjective by requiring the desired proportions of different types of question to be specified before analysing the exam, we think the procedure would then become unduly cumbersome.

Table 1 is a helpful checking procedure whenever any form of assessment is set. Table 3 is probably not so useful as an on-going tool. However, it provides a useful framework for exam setting which could easily be internalised and thereafter applied implicitly.

The biggest benefit is the surprise when the procedure is first used

Regardless of the limitations of Table 3, the procedure still has at least one big advantage. When applied to some previous computing exams, it became very clear that an inordinate amount of questions were simply of Type 1 — Memory of information. Even when the intention of the exam is to assess more deeply, it is easy to slip into simple memory questions because these are generally easier to write and easier to mark. I imagine that many examiners would find a similar surprise if the procedure were applied to their old exams. This provides good "shock therapy" to motivate us into taking greater care in future.

Once aware of the types of questions which assess deeper knowledge, it is perhaps not necessary to construct Table 3 for every exam.

Separating memory and non-memory questions

As already mentioned, it is our experience that many exams over-emphasise the memorisation of learnt information and fail to assess higher order reasoning skills. A useful indication of the extent to which a question (or an entire examination) relies on memory is to consider how much higher student marks are likely to be if the exam were open-book. This imbalance can be corrected by reducing the number of questions which simply require the recall or recognition of facts, definitions and procedure, and increasing the number of questions which fit into the other Types given in Figure 1.

In cases where assessing memorisation is important, it may be helpful to gather all the memory questions into one section of the exam. Other sections of the exam can then explicitly be designed to address non-memory-based issues and the examiner can more consciously choose question types which require deeper understanding. As mentioned above, we found that a successful approach in EUC was to assess the memory section via computer, and the non-memory section as a separate open-book exam.

Sometimes the need for broader changes are identified

When the use of the EWT identifies inadequacies in an exam, it will often be possible to address the inadequacies by adjusting individual questions, or replacing some questions with others. This can correct an imbalance in the percentage of marks allocated to each objective and it can also correct an inappropriate emphasis on testing certain forms of knowledge.

However, the use of this procedure may also identify deeper inadequacies which cannot so easily be corrected. It may be that the whole assessment of the course should be re-scrutinised rather than just the final exam. For example, are the course objectives clear? Are they assessable? What form of assessment is most appropriate for each of the objectives? If some of the objectives can be assessed through other forms of assessment prior to the final written exam, then this would allow the purpose, structure and content of the final exam to become clearer, and the overall assessment strategy more balanced.

The procedure may be applied to the complete assessment, not just the final examination

One of the primary purposes of the EWT is to highlight any significant discrepancy between the ideal and the actual relative allocation of marks in the exam in accordance with the various course objectives. However, the existence of such a discrepancy need not necessarily reflect negatively on the exam, since the final exam is not always meant to assess all of the course objectives. In particular, some of the objectives may have already been assessed using other instruments, and it may not be necessary to test them again. While it may be debated whether every objective needs to be assessed, it is certainly not the case that every objective needs to be assessed in the final examination.

There are two ways to address discrepancies which arise in such cases. First, one could reduce the list of objectives by selecting only those course objectives which are intended to be assessed in the final exam. Second, one could extend the scope of the whole procedure by applying it not just to the final exam, but to the combination of all forms of assessment within the course. This possible extension of scope is discussed further in the next section.

REFLECTIONS ON ASSESSMENT IN A UNIVERSITY COURSE

Up to this point we have centred our discussion largely around the setting of examinations because the focus of assessment at university is often on students' performance in a final, written task. However, as we suggested above, not all assessment must or should be conducted in this way. To place undue emphasis on a written, final examination is to underplay the potential role of assessment in tertiary education.

In practice, assessment lies at the very centre of university teaching and learning. Figure 6 shows the relationship between assessment, course objectives and what students actually learn. As we mentioned in the Introduction, the assessment given in a course largely determines what students will choose to learn. On the other hand, assessment is meant to provide a measure of what students have learned. From the instructor's side, the assessment that is given should be determined by the objectives of the course, while the outcomes of the assessment will give an indication of whether the course objectives have been met, i.e. whether students are able to perform in ways that the course was meant to equip them to perform. We believe that an awareness of this dynamic between teaching, assessment and student learning on the part of university teachers would result in improved student learning and better student performance in assessment tasks.

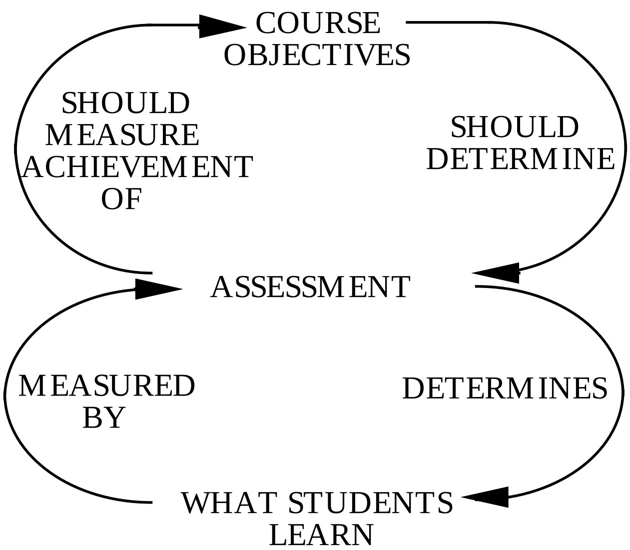

Figure 6: Inter-relationship between assessment, course objectives and student learning.

If we wish to get a balanced picture of a student's ability we need to use a number of methods of assessment, appropriately matched to what is being assessed. In the words of Graue (1993):

> "Choosing the appropriate assessment strategy is a validity concern; the tool must be relevant to the task at hand...No one strategy will fit all students, teachers, or schools perfectly, given the diversity in development, individual preferences, and cultural meanings.
>
> When we collect information from a variety of sources to address questions about a multi-faceted perspective on learning, we should expect that...interpretations of information should be made on patterns among information sources, avoiding heavily weighting single sources of information."

Moreover, not all assessment needs to be given at the end of a course. On the contrary, conducting some of the assessment while the course is in progress has several benefits. Firstly, it provides students with feedback which they can use to improve their future performance. Secondly, it provides the instructor with feedback about areas where students are having difficulties at a time when something can be done to remedy the situation. Thirdly, it is much easier to employ a variety of methods of assessment while the course is running than at the end. Since not all objectives are best assessed by means of a written exam question, the last point implies that it is more appropriate to assess certain objectives during a course, when other methods can be used[4]. For example, if we wish to assess a

---

[4] For ideas on other forms of assessment, see, for example Angelo & Cross (1993).

student's ability to set up an electric circuit given a circuit diagram and then use meters to take readings it makes sense to observe the student perform these tasks in the laboratory.

We recommend that for assessment to be a useful tool both for promoting and measuring student learning relative to a set of course objectives, the instructor should identify the appropriate methods of assessing each objective. The EWT provides a useful means of checking that an examination paper reflects the course goals in a balanced way, but if, as we suggest, other means of assessment are also used, then there needs to be an overall balance between the assessment given throughout the course and the course goals. When this is done, the instructor may decide to only test a subset of the course goals by means of a final written examination, and so fewer objectives will be included when the EWT is used. On the other hand, the principle behind the EWT can be applied to forms of assessment other than written examinations.

CONCLUSION

In this paper we have presented a tool for improving the quality of final examinations in two ways—by ensuring that the course objectives are adequately tested by the exam questions and that the questions appropriately assess a variety of types of knowledge, conceptual understanding and higher order cognitive skills. We have described how this tool was applied to two computer literacy courses, and have shown how the use of the tool led to changes in the assessment practices employed in those courses.

Our experience of taking a critical, in-depth look at final examination papers led us to consider the place of assessment in university science courses more generally. We have asserted that assessment lies at the centre of the teaching and learning process, and suggested how this central position can be exploited to improve the quality of student learning and of instruction. In addition to showing how final, written examinations can be improved, we have also suggested that they are only one means of assessment. Certain course objectives are better assessed using other methods. Moreover, formative assessment, that is, assessment given during a course, is beneficial both for guiding student learning and for giving feedback to the lecturer about the effectiveness of the course while it is in progress, i.e. while something can be done to modify it.

In conclusion, we maintain that by carefully planning the structure and timing of our assessment tasks we can improve the quality of student learning. The EWT provides one tool for assisting us in this task.

APPENDIX I: Worked example showing how to complete Table 1.

Question (1) assesses objectives a, b, c and is out of 5 marks
Question (2) assesses objectives a, d and is out of 10 marks
Question (3) assesses objectives b, e and is out of 20 marks

OBJECTIVE	MARKS	% (OF TOTAL 1)
a	15	20
b	25	33
c	5	7
d	10	13
e	20	27
TOTAL 1	75	100

APPENDIX II: Examples of questions types 1 to 5

There are two de Morgan's laws, both of which identify an equivalence between expressions involving disjunctions (the logical "or", represented by the symbol "$\lor$") and conjunctions (the logical "and", represented by the symbol "$\land$"). One law states that the negation (represented by the symbol "~") of a conjunction is equivalent to the disjunction of the negated conjuncts. In mathematical symbols, this law can be written:

$$\sim (A \land B) \ + \sim A \lor \sim B$$

An English example would be that claiming, "It is not possible to drive slowly and yet still get there on time" is the same as claiming "Either drive quickly or you won't get there on time". The other law is written symbolically as:

$$\sim (A \lor B) \ + \ \sim A \land \sim B$$

(notice that the disjunction and conjunction have been reversed from the first law). There is a standard method for proving equivalences such as de Morgan's Laws which makes use of truth tables. A truth table is a tool for establishing when a logical expression is true and when it is false. Suppose we construct a truth table for $\sim(A \land B)$ and another for $\sim A \lor \sim B$. These truth tables will show that whenever $\sim(A \land B)$ is true so is $\sim A \lor \sim B$ and whenever $\sim(A \land B)$ is false so is $\sim A \lor \sim B$. This constitutes a proof that the two expressions are logically equivalent.

The following questions all assess some aspect of truth tables, de Morgan's Laws and logical equivalence in accordance with the categories in Table 2.

Type 1a: Write down in symbolic form one of the two de Morgan's Laws.

The student is required to retrieve a memorised fact. Multiple-choice questions which require the recognition (rather than retrieval) of memorised information also fit into this category.

Type 1b: Describe the procedure for proving the equivalence of two logical expressions.

The student is required to write down a procedure which could not be deduced or derived, but must have been memorised.

Type 2a: Use any of the standard logical equivalences to rewrite the following expressions. In each case name the equivalences you use.

- $\sim(A \land B)$
- $A \land (B \lor C)$
- etc.

Each expression is one half of a standard equivalence; for instance the first is one of de Morgan's Law. The student is required to recognise which equivalences are applicable (in many cases there will be more than one), remember the other half of the equivalence and also remember the name of the equivalence.

Type 2b: Use any of the standard logical equivalences to simplify the expression $\sim(A \wedge (B \vee \sim A))$. Justify each step in the simplification by naming the equivalences you use.

> The question requires the use of de Morgan's Laws twice, along with several other equivalences. As in 2a, students are required to remember the names of a variety of logical forms such as "de Morgan's Laws", along with their symbolic representations. This question also tests whether the students' understanding is deep enough for them to recognise these logical forms embedded in complex expressions which have not been previously studied.

Type 3a: Prove de Morgan's Law $\sim(A \wedge B) + \sim A \vee \sim B$

> Students must recall the procedure for proving equivalences via truth tables and correctly apply the procedure to this known equivalence. It is quite possible that they have seen this exact proof either in a lecture or as a tutorial question.

Type 3b: Prove whether or not the expression $\sim(A \wedge B \wedge C)$ is equivalent to $\sim A \vee \sim B \vee \sim C$.

> This is a natural generalisation of de Morgan's Law, but one which students will not have seen before. The procedure for proving the equivalence is the same as they have encountered before.

Type 4a: Given de Morgan's Law $\sim(A \wedge B) + \sim A \vee \sim B$, rephrase the following sentence using the word "or" in a way which does not change its logical meaning — "It is not possible to drive slowly and yet still get there on time."

> This question gives students the logical structure which they must follow, and presents them with an English statement which closely resembles one which they have seen before.

Type 5a: Two logical expressions are equivalent if they are true under exactly the same circumstances. Construct a truth table for the expression $\sim(A \wedge B)$ and another for the expression $\sim A \vee \sim B$. Are these expressions equivalent? Justify your answer.

> The logical expressions are ones which the student has seen before and the question dictates the structure of the required answer. The answer will be identical to the answer for the question in Type 3a above, but it requires much less to be remembered by the student (though they must still remember how to construct truth tables).

Type 4b: As for 2b above, but with a more complex expression to be simplified and a list of the standard equivalences provided.

> Providing a list of required facts and formulae (e.g. by making the exam open-book, or by including a list as an appendix to the exam) reduces the difficulty of a question like this. However, it also means that a more complex context can be specified which tests the ability to apply the standard equivalences more extensively. In 2b for instance, the simplification may only take a few lines and a lot of the marks will be allocated to the correct memorisation of the equivalence names. But in 4b, the simplification may require much more work and the majority of the marks can be allocated to the accurate application of the equivalences and to the overall structure of the simplification strategy.

Type 5b: Two logical expressions are equivalent if they are true under exactly the same circumstances. Construct a truth table for the expression $\sim(A \wedge B \wedge C)$ and another for the expression $\sim A \vee \sim B \vee \sim C$. Are these expressions equivalent? Justify your answer.

The format of this question applies the procedure given in 5a to the new context described in 3b.

APPENDIX III: Examples of question types 6 to 10

Type 6: Two students are in the supermarket. One student takes a tin of fish off the shelf and a packet of biscuits. Say which, if either, of the students is correct and why:

> Student #1: The tin of fish feels colder than the packet of biscuits. It has a lower temperature.
>
> Student #2: No you are wrong. Both the fish and the biscuits have been sitting on the shelf for quite some time, so they must both have the same temperature.

This question tests students' conceptual understanding of temperature. Students with a good conceptual understanding would know that the objects in the supermarket would all be in thermal equilibrium with each other, but that objects feel different to the touch because of different thermal conductivities.

Type 7: Suppose an object is hanging on a spring and bobbing up and down. Some experiments are performed to find out whether the length of the spring and the mass of the object influences the bobbing rate. The results of three experiments are shown in the following table (McDermott, 1996, p. 140).

	Exp 1	Exp 2	Exp 3
Length of spring	12 cm	27 cm	12 cm
Mass	50 g	72 g	72 g
Bobs in 10 sec	54	30	45

A. Does length of the string *influence* the bobbing rate? Tell which two experiments should be compared to answer this question and explain your reasoning.

B. Does the length of the string *determine* the bobbing rate? Tell which two experiments should be compared to answer this question and explain your reasoning.

This question is testing whether students understand the difference between the reasoning involved in figuring out if something influences an experimental result as opposed to determining it. No content knowledge is assumed.

Type 8: A manufacturer's catalogue states that the 2 kW kettle they sell will boil 2 litres of water in 5 minutes and for a cost of less than 1/2 cent. Given that electricity costs 3 cents per kW h, check the accuracy of each of these claims. (Assume initial water temperature of 20^{o}C; specific heat capacity of water is 4180 J/kg K).

This question requires students to integrate knowledge learnt while studying the topic of heat with knowledge acquired during the study of electricity. Often these topics are encountered several months apart in an introductory physics course.

Type 9: Suppose two blocks of the same metal are heated. The first block receives 490 joules of heat and rises 18 ^{o}C in temperature. The second block receives 280 joules of heat and rises 36 °C in temperature. If the mass of the first block is 35 g, what is the mass of the second block? Solve using dimensionless ratios. Show all your working.

> This question tests whether a student can apply the technique of using dimensionless ratios to solve a numerical problem in which insufficient information is available to solve the problem using equations. Some content knowledge is required, namely the knowledge that the heat transferred is directly proportional to the temperature change of an object and the mass of an object.

Type 10: Describe the evidence leading to the conclusion that the planes of the ecliptic and the equator do not coincide and describe what we would observe if they did coincide (Arons, 1977, p.168).

> This question tests whether students can go through the reasoning process required to make inferences about something which they cannot observe directly from something which they can observe directly. This process is represents a central part of the epistemology of physics.

REFERENCES

Angelo, T. A. & Cross, K.P. (1993). *Classroom Assessment Techniques. A Handbook for College Teachers. Second Edition.* San Francisco: Jossey-Bass.

Arons, A. (1977). *The Various Language*. New York: Oxford University Press.

Graue, M. E. (1993). Integrating theory and practice through instructional assessment. *Educational Assessment* **1**(4), 283-309.

Habeshaw, S., Gibbs, G. & Habeshaw, T. (1993). *53 Interesting Ways to Assess Your Students*. Melksham, UK: Cromwell Press.

Heywood, J. (1989). *Assessment in Higher Education (Second Ed.)*. Chichester: John Wiley and Sons.

McDermott, L.C. (1996). *Physics by Inquiry,* Vol. I. New York: Wiley.

Ramsden, P. (1993). *Learning to Teach in Higher Education*. London: Routledge.