Supporting Good Practice in Assessment in Mathematics, Statistics and Operational Research

Briefings • Guide for Heads of Departments • Guide for Lecturers
(a guide for students is also available)

Neil Challis, Ken Houston and David Stirling

The Higher Education Academy Maths, Stats & OR Network

Resources to support the learning and teaching of Mathematics, Statistics and Operational Research in Higher Education
Maths, Stats & OR Network, The University of Birmingham, Edgbaston, Birmingham, B15 2TT.
Phone: +44 121 414 7095 • Email: info@mathstore.ac.uk • Web: www.mathstore.ac.uk

Supporting Good Practice in Assessment in Mathematics, Statistics and Operational Research

Briefings

Guide for Heads of Departments

Guide for Lecturers

(a guide for students is also available)

Neil Challis, Sheffield Hallam University

Ken Houston, University of Ulster

David Stirling, The University of Reading

Case Studies contributed by

Cliff Beevers, Heriot-Watt University

Joe Ward, University of Loughborough

Supporting Good Practice in Assessment

Briefings and Guides

We hope you find this series of guides and briefings on supporting good practice in assessment in Mathematics, Statistics and Operational Research to be useful. The series owes much, and draws heavily (with permission and indeed encouragement) from the Learning and Teaching Support Network Generic Centre's series of Assessment Guides and Briefings.

Our aim has been to convey many of the ideas therein, but to interpret and add to them in such a way as to bring out their relevance and usefulness in the context of Mathematics, Statistics and Operational Research (MSOR).

There are six briefings including case studies, and two guides. A separate guide for students is also available. We hope and envisage that the guides and briefings may be useful in a variety of ways: in supporting initial professional development for new lecturers; in supporting continuing professional development for existing lecturers; in supporting both interested individuals and those with relevant responsibilities such as oversight of assessment in an MSOR grouping.

The series has been created and edited by Neil Challis, Ken Houston and David Stirling, with additional case study material contributed by Cliff Beevers and Joe Ward. Our thanks go to all the authors of the LTSN Generic Centre's series upon whose work we have drawn; and also to the LTSN Maths, Stats & OR Network for their financial support for the project which created these materials.

We welcome your feedback on these materials, particularly any information about where they have proved useful, and also ways in which they could be improved.

Neil Challis, Sheffield Hallam University
Ken Houston, University of Ulster
David Stirling, The University of Reading

The Editors

Neil Challis is Head of Mathematics at Sheffield Hallam University. He is a former secretary of HoDoMS (Heads of Department of Mathematical Sciences). He is interested in the ways in which mathematics is learnt, taught and assessed, and particularly in the impact of technology upon those, and has published a number of articles about these topics.

Ken Houston is Professor of Mathematical Studies in the School of Computing and Mathematics at the University of Ulster. He was Head of Department of Mathematics at UU for 10 years. He is a former chairman of HoDoMS (Heads of Department of Mathematical Sciences), of UMTC (Undergraduate Mathematics Teaching Conference), and of ICTMA (International Conference on the Teaching of Mathematical Modelling and Applications). He has published a number of articles and books on the teaching of mathematics at university level.

David Stirling is at The University of Reading where he was Head of the Department of Mathematics until 2003. He is a mathematical analyst by trade and the author of books on Mathematics Analysis and Integral Equations. He is currently partly seconded to the Faculties of Life Sciences and of Science as Director of Teaching and Learning, in which role he chairs the University's Committee on Examinations and Assessment.

The Other Contributors

Professor Cliff Beevers has directed the CALM Project for Computer Aided Learning in Mathematics since 1985 and has taught at all levels at the Heriot-Watt University over a career spanning more than three decades. He is currently co-director of the Scottish Centre for Research into On-Line Learning and Assessment (http://www.scrolla.ac.uk). SCROLLA is collaborating with all the leading Scottish educational agencies to pilot the automatic delivery of the National Assessment Bank of questions in Higher and Advanced Higher Mathematics directly into schools. Cliff has recently been awarded the OBE for his work in Higher Education.

Joe Ward is Senior Lecturer in the Department of Mathematical Sciences at Loughborough University. He has published a number of articles and books on the teaching of mathematics at university level. He is presently engaged with the FDTL4 project HELM (Helping Engineers Learn Mathematics).

Briefings

Guide for Heads of Departments

Guide for Lecturers

Assessment in MSOR:

Briefing Number 1: Key Concepts

Introduction

This briefing draws heavily on the corresponding booklet in the Generic Series, Booklet 7, written by Peter Knight (Knight 2001), and attempts to relate these concepts to MSOR.

In the past mathematicians have been slow to learn and to use the language of education. Today we recognise our need to be professional educators as well as professional mathematicians, and this chapter introduces some of the key educational concepts relating to assessment. Many of these concepts are already part of our culture, even if we do not use the recognised educational terms to identify them.

In the past mathematicians have been slow to learn and to use the language of education

It is not hard to do this and it helps us if we know the language when we engage in discussion with generic educators or when we read educational research. This chapter should be of particular value to those new to our profession.

The *Guide for Lecturers* at the end of this book draws attention to the need to align assessment tasks, assessment methods, methods of learning and intended learning outcomes of programmes and modules. (A *programme* is the whole course that a student takes and a *module* is one of the building blocks of a programme. In some cultures the word *course* is used to describe a programme, and in others it describes a module. We have tried to distinguish carefully between these by avoiding the ambiguous use of the word *course*.) The *Guide* thus advocates giving careful thought to the purposes of assessment and to the selection of appropriate assessment methods and assessment tasks. The *Key Concepts* this briefing describes are *Summative Assessment* and *Norm Referencing*, which relate to the purposes of assessment in regard to reporting, and *Formative Assessment* and *Criterion Referencing* which relate to the purposes of assessment in regard to student learning. The briefing also looks at the concepts of *Reliability*, *Validity*, *Affordability* and *Usability*.

Summative assessment and formative assessment

All MSOR lecturers, including recent graduates and others new to the profession, would (probably) have experienced both summative and formative assessment when they were students, but may not have appreciated the distinct purposes and values of each.

Summative Assessment is the term used to describe any assessment procedures which "sum up" a student's performance in the programme or module. Summative assessment attempts to measure the extent to which a student has achieved the learning outcomes of the programme or module, and this measure of achievement is then reported to the student, to the awarding body and, usually in some condensed form, to the world. A student may indeed learn mathematics from having knowledge of this measure, but more likely, the knowledge will merely produce a feeling of euphoria, satisfaction or dissatisfaction, depending on where the measure falls on the scale from first class through pass to fail. It is not unknown for some students to take a strategic approach to their learning and to assessment tasks so that they learn and do what they think is just enough to achieve a particular measure, say "pass".

Summative assessment attempts to measure the extent to which a student has achieved the learning outcomes

Formative Assessment is the term used to describe any assessment procedures which "form" or mould a student's learning and which "inform" him or her, not only of the measure achieved, but also of the gaps in their knowledge, the inadequacies of their performance, and then, knowing these, how they may do better. Detailed feedback should always be given to students on their performance on formative assessment tasks: "Yes, this answer is good because…; no, that answer is inadequate because…., and here is how you could have answered." Formative assessment tasks should include some tasks that give a student the opportunity to demonstrate what they know and can do. They should also include some tasks that are challenging, which even the best students may not be able to complete and which will thus provide opportunities for further learning after feedback. Such tasks might be considered to be unfair were they to be included in summative assessment procedures.

The traditional unseen, written, three-hour examination is the most favoured summative assessment task within the MSOR community. Traditionally students never see their answer books again after they are handed in. They will never be told which answers are correct and which are not. They may never know what percentage mark they scored in a module, nor even whether they passed it or not. All they may receive is the overall level of award – "upper second" – or, in earlier years, "permission to proceed". These traditions are changing. Many students today are given transcripts of the marks achieved on each module. They are then able to apply a published algorithm to satisfy themselves that the level of award matches the module performances. Some institutions are returning marked scripts, or at least providing an opportunity to view them, so that students who feel aggrieved with their award may harbour no lingering doubts about the fairness of their

treatment. A by-product of this is the formative effect achieved by having mistakes highlighted.

The traditional unseen, written, three-hour examination is the most favoured summative assessment task within the MSOR community

Quite often in the MSOR community the same assessment tasks are used for both formative and summative purposes. Thus homeworks and class tests are used in these ways – to test student learning for the purpose of reporting and to give feedback for the purpose of enhancing further learning. However experience is showing that this dual role may not always be best. When students are given a grade or score for a piece of coursework, they are inclined merely to feel content or discontent depending on the score, and then to file the work away without looking at it carefully to learn from any failures. On the other hand, if a grade or score is not given, but a careful commentary is provided, then the student is inclined to read the commentary and to learn from it. See Black and Wiliam (1998), Onwuegbuzie and Leech (2003) and Taras (2002)

Another aspect that the MSOR community might begin to consider is the idea that a particular assessment task be scored either "0" or "1", with a "1" being required in each task to "pass" a module. A "1" would be awarded if the task is completed satisfactorily (and "satisfactorily" should be defined), but a percentage score for inclusion in the overall module score is not assigned. Again the emphasis would be on giving feedback to enhance learning. The task is for formative assessment, but there is a strong requirement to work at the task and to submit at least a satisfactory performance. This would obviate the need to decide on (and defend) the percentage contribution a particular task would make to the overall module score.

Norm referencing and criterion referencing

The main purpose of a norm referenced assessment scheme is to place students in rank order. To achieve this, assessment tasks that are discriminating should be set, that is, the assessment should produce a good spread of marks for the students in the class. The examiners must have some expectation of the proportions of students in the different grades or levels of award that are reasonable. Thus, in final year, the examiners might say that all students above the median should get an upper second with the top 10% (say) getting a first, while all below the median would get a lower second except for some percentage at the bottom who would get a third. A very few really hopeless cases would fail. The assumption behind this procedure is that the distribution of abilities in the student population does not vary from year to year, and this might be a reasonable assumption to make when there are many students in the class. It also nullifies any variability in the standard of the examination papers and of the teaching from year to year. Norm referencing is thus a way of treating marks rather than a system for awarding them.

On the other hand, criterion referencing is a system for awarding marks and it is open to the possibilities that everyone may fail or everyone may get a first. In this system, the examiners decide beforehand what counts as a good performance, giving as precise as possible "level descriptors" for each of the different levels or grades of award that they wish to make. They then judge the evidence of achievement from each student against these level descriptors.

The MSOR community usually uses a mixture of these modes of assessment. Thus in treating the raw marks from an aggregation of examination marks, the examiners might say that students who score 70% or more will get a first, and so on. This is criterion referencing. But then they might look at the proportions of students in each class and, if these are not quite what they were expecting, without changing the rank order, the examiners may make adjustments to the class boundaries. This is norm referencing. On the whole this system has worked fairly well, but it is more difficult to operate in a situation where students are given marks and can inspect their papers.

Also in the marking of project work, criterion referencing is widely used. But then problems arise when comparing the achievements of different students. The level of difficulty of projects may be different and this is hard to detect when each student only does one project and they all do different ones. The assumption has to be that all projects are equally difficult, that all examiners mark to the same standard and all supervisors are equally good (or bad).

Norm referencing is thus a way of treating marks rather than a system for awarding them

Reliability

Reliable assessment measures are objective, accurate, repeatable and correctly marked and recorded. They are objective in the sense that there is no bias between examiners. They are accurate in the sense that the methods used are stable over time and sensitive to small differences between students. They are repeatable in the sense that they are consistent from case to case, over time and across the student cohort. They must be correctly marked, correctly recorded and the correct aggregating algorithms must be applied correctly.

The intention is to produce objective data about student performance and the MSOR community comes very close to this through its use of unseen, timed, written examinations. Questions are usually phrased precisely and unambiguously, and a very detailed and tight marking scheme is prepared. Alternative, valid solutions are considered and equitable marking schemes are written. With large classes, the scheme will be piloted and modified if that is found to be necessary. When more than one examiner is employed to mark the examinations, they will each mark samples and discuss their experiences in interpreting the scheme. They will agree

modifications to apply to the scheme. This eliminates bias between examiners to a considerable extent.

The intention is to produce objective data about student performance

It also helps each examiner to be consistent across the whole of his or her cohort. Sometimes it is useful to mark the questions one at a time right through the class. Then it is easier to remember how marks are to be awarded and there is less "drift" through the cohort. In an ideal world, where the same syllabus is being examined from year to year, succeeding examination papers should be producing similar measures of achievement for similar students.

However, assessment is a complex human endeavour, and will never be completely reliable. But the MSOR community is pretty good!

Validity

Written examinations are pretty *reliable*, but they only have limited *validity*. To be "valid", assessment must assess all the things we expect students to learn which are considered to be of value. "Demonstrate proficiency in oral communication", may be a learning outcome, but it cannot possibly be assessed by written examination. Similarly for "demonstrate effective teamworking skills". "Carry out and report on an extended investigation" has to be assessed in a different way – usually a Project. As the *Guide for Lecturers* points out later, there has to be alignment between learning outcomes, methods of learning, assessment methods and assessment tasks. The more complex the task, the less likely it will be that a written examination will be a valid means of assessment; the more we simplify the assessment, the more reliable it becomes, but at the expense of validity.

However, it is possible to get acceptable levels of validity when assessing complex authentic behaviour, but at a cost!

This is costly and complicated

Affordability

Assume that a definition of some complex attribute such as "presentation skills" can be written and assume that a description of what could count as good evidence can be agreed, then, with enough resources reliability can be improved. It requires multiple observations of performance by several trained observers, and the performances should be fair samples of the range of possible performances. This is costly and complicated. Assessors have to be given the time and have to be trained. Good grade indicators have to be developed and tested, and second or even third markers have to be provided, each of whom will have to make multiple observations. The scheduling is complicated. Izard (1997) and Izard et al (2003) give suggestions as to how this might be achieved in MSOR. Haines and Houston (2001) present a case study which illustrates how training and testing might be achieved in MSOR.

While it may not be possible, due to resource constraints, to achieve this level of reliability for the assessment of complex behaviour in just one module, it may be possible to do so over an entire programme.

Usability

This is concerned with the *usefulness* of assessment to each of the internal stakeholders – students, lecturers, managers, and each of the external stakeholders – postgraduate schools and employers. What is it each wants to know? For students, formative assessment is useful for learning and summative assessment for progression and award (or not). Lecturers need information about the progress of their students and managers are keen to observe good quality. Postgraduate schools will want considerable detail of a student's whole profile and employers will want assurance that the student knows the subject, perhaps even specialised topics, and whether the student has good skills. None of this is particular to MSOR, but it is important for the MSOR community to think about these things and to contribute to institutional debates when policies are being drafted.

References

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

Haines C and Houston K (2001) "Assessing Student Project Work" in Holton D (editor), *The Teaching and Learning of Mathematics at University Level*, Dordrecht: Kluger Academic Publishers, pp 431-442.

Izard J (1997) "Assessment of Complex Behaviour" in Houston SK, Blum W, Huntley I and Neill N, (editors) *Teaching and Learning Mathematical Modelling*, Chichester: Albion Publishing, pp 109-124.

Izard J, Haines, C, Crouch R, Houston K and Neill N (2003) "Assessing the Impact of Teaching Mathematical Modelling" in Lamon SJ, Parker WA and Houston SK (editors) *Mathematical Modelling: A Way of Life*, Chichester: Horwood Publishing, pp 165-177.

Knight P (2001) *A Briefing on Key Concepts*, LTSN Centre Assessment Series No 7, York: LTSN Generic Centre.

Onwuegbuzie AJ and Leech NL (2003) "Assessment in Statistics Courses: more than a tool for evaluation"**,** *Assessment and Evaluation in Higher Education*, **28** (2), pp 115-127.

Taras M (2002) "Using Assessment for Learning and Learning for Assessment", *Assessment and Evaluation in Higher Education*, **27** (6), pp 501-510.

Assessment in MSOR:

Briefing Number 2: Assessing Large Groups

Introduction

In this briefing, which draws on the corresponding booklet in the Generic Series, Booklet 12, written by Chris Rust (Rust, 2001), we briefly draw attention to a range of ideas and issues in dealing with assessment of large groups, and then present two extended MSOR case studies where large groups of students are being assessed in interesting ways, particularly exploiting the possibilities of computer-aided assessment (CAA).

To summarise Rust, it is recognised that assessment strategy is intimately connected to student learning. Group size can affect how much assessment can be done. If, because of a large group, assessment is carried out less often (simply not enough time) or less well (less rigorously, with more superficial feedback, taking longer to be returned), then this can have a negative effect on student learning and achievement. We have expressed throughout this booklet the view that what you assess is what you get. If you assess less, then do you get less?

assessment strategy is intimately connected to student learning

Joe Ward, in Case Study 2 in this briefing, criticises Rust's generic guide: "Rust's article was essentially directed at the non-scientific community". While we believe this might be a widely held view, it behoves us as the MSOR community to be prepared to consider what we might learn from other discipline communities who face similar problems to ourselves in dealing with large groups. For example students of MSOR may work on discursive tasks such as project reports, and so we may have something to learn from more discursive disciplines. For that reason we summarise some more generic thoughts going beyond CAA before proceeding to present the case studies. To those whose interest is caught by any of these thoughts, we do commend the Rust booklet for more details.

Strategic issues

Rust (2001) raises several possible strategic suggestions: one example is a wholesale move to CAA at first year level; another is to revise regulations, for instance streamlining resits, or relaxing restrictive conditions designed for less populated days. An interesting suggestion is to use "course requirements", where for instance it may be a requirement for students simply to have completed a piece of work before being allowed to sit an exam, although the work is not summatively assessed.

Creative solutions

Rust (2001) presents creative suggestions under six headings.

Front-ending

Marking can be simplified by improving how the students are set up for their assessment. This can involve providing very full briefing instructions, and anticipating what students may do wrong or misunderstand about the task. It may involve clarifying the criteria by which the assessment will be marked. It is interesting to note here the difficulty of getting students to engage with this: one way is to get students to mark past students' pieces of work. See Nelson (1994) for details.

Doing it in class

The heading says it all: what can you do in allocated class time? Rust (2001) presents three suggestions:

- Give general feedback. This can be followed up with students discussing in class which bits of the feedback applied to them.
- Set assignments which can be marked in class. This could involve oral or poster presentations, also contributing to development of communication skills; or peer assessment, perhaps marking each others' tests.
- Set assignments which can be undertaken in class. Students may do an in-class test, or perhaps have to write up and submit a modelling report within a session, which will necessarily be briefer than it otherwise might be.

Self- peer- and group-assessment

These are discussed in Briefing number 6 in this booklet.

Mechanising the assessment

We present below two examples of practice in the MSOR community as case studies. CAA (computer-aided assessment) is of widespread interest, and many current developments in MSOR may be found in the online journal edited by Beevers (2003).

Strategic reduction

One way to deal with too much marking is to reduce the amount of assessment, but there are real dangers in doing this. Many students now work strategically, responding only to the pressure of assessment. If that assessment is reduced, they may not devote the time they need to absorb and digest ideas until it is too late.

One way to deal with too much marking is to reduce the amount of assessment, but there are real dangers in doing this

It is important if reducing assessment load to do so strategically, and not in an ad hoc manner. Rust (2001) suggests reducing repetitive assessment, varying assessment

methods; tailoring methods to the learning outcomes; noting that first year assessment, being mainly about progression, does a different job from that in later years where classification becomes an issue; and using course requirements as mentioned earlier rather than marking everything.

More efficient feedback

Rust also suggests making feedback more efficient, for instance by: giving general feedback to the whole class; using feedback forms, which if electronic may involve using statement banks or model answers; and using CAA which gives some feedback.

So to the two MSOR case studies.

Case Study 1: Teaching Service Mathematics to Large Classes

C E Beevers (Heriot Watt University)

This case study describes the strategy employed in the teaching of Service Mathematics to large groups of first year undergraduates at a typical Scottish university over the last two decades. In particular, it outlines the role of the computer in enhancing student motivation. It sets in context the features of an automatic assessment system that has delivered both formative and summative computerised testing to thousands of undergraduates in science and engineering through projects like CALM and MathWise (www.calm.hw.ac.uk and www.bham.ac.uk/mathwise), SUMSMAN (see the earlier work described by Beevers et al (1996)) and SCHOLAR (scholar.hw.ac.uk).

Introduction

This short article draws on the experiences of the author in the teaching of first year Calculus to large groups of engineering and science undergraduates in a first year university course at the Heriot-Watt University in Edinburgh. Recently, I retired from full-time teaching but, for almost twenty years, I had taught the large Service Mathematics class with numbers in excess of 200 students per year. This case study seeks to describe the lessons learnt in teaching such a large class and the role that the computer can play in such delivery.

The motivating examples were fun to create…

Early days

The CALM Project for Computer Aided Learning in Mathematics started in 1985. Originally CALM set up a series of weekly computerised tutorials to support the teaching of a first course on Calculus to large groups of engineering and science undergraduates. The resources in the computerised tutorial were created using the language Pascal and replaced the traditional pen and paper approach that was struggling to provide sufficient support for the large numbers of undergraduates. Each week throughout a 25 week course mathematical topics were covered to the formula:

summary of the theory delivered in lectures;

worked examples to illustrate method;

motivating examples to bring the subject to life; and

test section to allow students to assess their own progress and enable teachers to monitor class performance.

The motivating examples were fun to create and did engage the students in problems like filling a cylindrical beaker from a conical dispensing coffee machine, calculating the escape speeds from a variety of planets, maximising the trajectory of an inter-continental ballistic missile and taking the largest plank around a corner in a tunnel in the game known as Escape from Colditz. However, from the outset the students chose to spend much of their time practising within the test section. Quickly the tests acquired a number of features to cater for student and teacher needs such as:

questions that required mathematical expressions as answers rather than the simpler multiple choice format;

multi-staged questions;

the delivery of each test in a variety of feedback modes spanning no help on screen to visible ticks and crosses and the ability to reveal an answer if stuck;

levels of question from "Very Easy" to "Hard" with the choice in the student's hands; and

the use of random parameters within questions so that each time the test was taken different numbers appeared thus maintaining the freshness of each example.

The formative assessment offered by CALM was welcomed by the students

The formative assessment offered by CALM was welcomed by the students and one of them spoke for many when he reported "The CALM tests and their feedback gave me the first indications that I was coping with my university course." In addition, there was a crude but effective reporting tool that gathered up the student test records at the end of each week ready for human scrutiny. The lecturer, then, had to go through manually to see how each individual was progressing. This was time-consuming though worthwhile, as it meant that comments could be relayed back to the students at the start of the next week. Indeed, Dr J H Renshaw (now of Southampton University) constructed a reporting system, which allowed the lecturer to view the class records and then visit the file of any individual student, see their marks and review their answers to any of the questions. The lecturer could report back to students by the creation of text messages which when stored on the file server displayed next time the student logged onto the system. This became a powerful way of

communicating with the large numbers of students using the system and provides an early example of a database results system with reporting capability. It should be noted by younger readers that in the middle 1980s email was but a twinkle in the electronic eye!

CALM was delivered over a couple of decades to thousands of students and from the late 1980s this included a diagnostic test at the start of the course. This diagnostic test was based on an earlier pen and paper multiple-choice test designed at Glasgow University. There were 25 questions covering arithmetic, simple algebra, trigonometry, more complex algebra and a couple of questions on geometry based on knowledge of the Scottish Higher Mathematics syllabus (similar in level to the AS award in England). Written in Authorware, the diagnostic test provided the students with instant feedback on which areas they needed to revise. Lecturers on the course then gave the students some refresher lectures based on the topics in the diagnostic test.

Another strategy used with some effect grew out of a 3-day revision course in 1988. The formula was as follows:

before the course began students were asked to pick out a number of topics on which they wanted further explanation;

conventional one hour presentations were then prepared to cover these problem topics, one at a time;

at the end of the hour-long lectures the students were again asked to write down what they still did not understand;

the students were then directed to relevant sections of the CALM software for one more hour to try to eliminate their problems;

there then followed a further request for the students to say what was still troubling them; and

finally, a round-up session was held to attempt to resolve any remaining difficulties delivered in a mixed style between tutorial and lecture.

This recipe sought to produce a decreasing list of problem topics for the students. The revision course worked well in this way and the interested reader is directed to Beevers et al (1989) for fuller details. However, this condensed formula working over a few days would not easily translate onto a weekly or termly timetable. I was determined, though, to exploit this successful approach, which discovered and then reduced student problems as the course itself progressed through term. The solution adopted was as follows:

give the conventional lectures as usual each week;

run the computerised CALM tutorial to back these up; and

then, ask the students what if anything remained as a difficulty for them at the end of the week.

the students learned to "own" the extra sessions

In the middle of the next week an extra session was organised at which any problem topics were re-considered as a mixture of worked solution on the blackboard and tutorial format with the students trying examples for themselves. The extra session would not run if the students did not come forward with problem topics. In this way the students learned to "own" the extra sessions and over the years these extra sessions worked well with perhaps 5% - 10% of the class taking advantage of this optional extra session. It was made clear from the start that this was not an additional class to go over the hardest questions on the tutorial sheet but rather a second chance to learn. From that revision course in the late 1980s the extra session ran each week with perhaps two omissions over the next fifteen years.

The Mathwise years

In 1992 CALM became one of the three principal resource centres for the distributed TLTP (Teaching and Learning Technology Programme) project known as Mathwise. The Heriot-Watt remit was to support authors using Authorware and supply advice on assessment. A formative assessment policy developed in which the learning resources of Mathwise were littered with examples for students to try but with no record kept of how they had performed. This self-testing approach was designed to build up student confidence. At the end of some of the more popular modules a more formalised test was constructed in which students were encouraged to try questions but this time a record of achievement was stored on the system. In this way the lecturer could monitor individual or class performance. This gave some measure of continuous assessment for some of the Mathwise modules.

As part of the Mathwise Project the assessment engine was employed to grade performance in the end of term test for Service Mathematics for first year students at Heriot-Watt University. It was introduced following a proper educational evaluation in 1994 and from 1995 provided part of the student mark for the course (see Beevers et al, (1995) for a detailed discussion on the evaluation). One bug in this testing approach was the inability of a student to return to a question once it had been completed, unimportant in formative assessment but vital in summative testing. Another issue flagged by the educational research was the need for partial credit.

Another issue flagged by the educational research was the need for partial credit

The Mathwise design of questions had allowed for up to six prompts in a question with each one being able to be split in two at the press of a steps button. The steps button had been introduced to help students who were stuck and needed extra help. However, the same device opened up one way of providing for partial credit in summative tests and there is more detail on this in Beevers et al (1999) and McGuire et al (2002). Despite the large numbers involved in automatic summative testing at Heriot-Watt it has always been the practice to emulate the security surrounding the more usual paper examinations. So, although this requires

additional work, each student has to come along at a given time, prove their identity, sign an examination agreement slip and follow all the usual examination regulations.

As part of the SUMSMAN Project Paisley University and Napier University also used the end of module tests in Mathwise to grade, in part, their students. The randomisation feature, highlighted earlier, remained a critical part of this whole process.

Enhancing the assessment systems

CALM then produced a series of commercial CDs called Interactive PastPapers in collaboration with Lander Software of Glasgow (Interactive PastPapers for A Level/Higher Mathematics, Lander Educational Software, Glasgow (1997)). IPP, as it became affectionately known, built on the features from CALM and Mathwise and included in addition:

a dynamic Input tool to help with one-line entry of mathematical answers;

marking of multiple entries in the form of ordered and unordered lists of responses to cater for answers like vectors and factors respectively;

more precise numerical routines to aid the checking of numbers to given significant figures or decimal places; and

additional flexibility in question choice and answer layout.

Moreover, it was possible in IPP to return to a question and amend it paving the way for more effective summative testing which was introduced into the University in the late 1990s.

In 1997 the CALM team decided to move away from the restrictions of Authorware and create the entire assessment package within the Delphi environment. This development looked ahead to the delivery over the web. It became prudent to make strategic alliances and with the University of Cambridge Local Examination Syndicate (UCLES) and the commercial firm EQL of Livingstone in West Lothian suitable partners were found. CALM, UCLES and EQL all contributed to the new assessment engine called CUE. CUE's design allowed for new answer types to be created. So, as well as the mathematical expression checker, multiple choice, multiple response, multiple hotspots and word match part types soon emerged in the new engine. The role of optional steps became clearer following further trials of the software with undergraduates and in secondary schools (see Fiddes et al (2002) and McGuire et al (2002) for reports on these findings). Throughout these developments the various assessment systems were used to enhance the learning for the undergraduates.

At the start of the year 2000 the SCHOLAR Project at Heriot-Watt University began. SCHOLAR provides notes and interactive sequences on-line creating learning and assessment resources in print and on the web to cover the Higher and Advanced Higher secondary school syllabuses in Biology, Chemistry, Computing, Mathematics and Physics. SCHOLAR is currently delivering to all 32 Scottish Education Authorities and to the independent schools sector North of the Border. SCHOLAR is an example of a Virtual Learning Environment with features like a web board for communication, a Frequently Asked Question section and a reporting system on student performance supplementing the learning materials. CUE is at the heart of SCHOLAR and is providing interactive assessment over the web in a variety of subject areas. As I write this article an A and AS level version of SCHOLAR is shortly to be piloted in some 100 English schools.

Conclusions

The experiences through all these projects has led to the analysis that computer aided assessment has a number of roles to play within the learning environment (as discussed by Beevers and Paterson (2002)) and summarised as follows:

Diagnostic testing with its emphasis on alerting students to their strengths and weaknesses at the start of a course of study;

Self-testing during which students are encouraged to practice their knowledge, understanding and skills without any attempt to gather marks or record their results;

Continuous assessment with the balance subtly moving to the recording of marks and responses so that the teacher can be informed and contribute to the feedback;

Summative testing in which the grades count towards some award.

computer aided assessment has a number of roles to play within the learning environment

Automatic delivery of diagnostic tests worked well in CALM, self-testing is a feature of Mathwise and SCHOLAR and electronic continuous assessment features in CALM and Mathwise. The role of the computer in grading testing remains a subject of intense research through projects like PASS-IT (see http://www.pass-it.org.uk). However, marking of student performance by computer is becoming more popular and it remains important that this measure of achievement receives more scrutiny in the years ahead.

Case Study 2: Assessment of Large Groups

J. P. Ward (Loughborough University)

Introduction

Having dealt with large groups of students (typically 100 but up to 200) all my academic life I read with interest the report by Chris Rust (Rust, 2001). Many of the assessment strategies discussed there referred to discursive subjects commonly found in the Arts and Humanities. Of the six approaches outlined I can only realistically regard the assessment strategy of *doing it in class* as being of any use in scientific disciplines. But even here valuable lecture time is wasted if timetabled slots are used for examination purposes. Very little was said on Computer Aided Assessment (CAA). Not surprisingly *this* form of assessment can be used to play a significant role in scientific disciplines. Though Rust's article was essentially directed at the non-scientific community I wholeheartedly agree with his view that assessment strategies should be at the heart of learning and teaching strategies.

At Loughborough University, in the delivery of mathematics to engineers, we use two forms of assessment: coursework (given throughout the Semester) and the traditional end-of-semester written examination. We used the coursework element to ensure that all students retain interest and focus throughout the twelve week module. We experienced (no doubt in common with many other institutions) that if assessment was through a single timed written paper then many students switched off until two or three weeks before the exam by which time it was too late to recover. Ideally, each student would be given an individual coursework assignment distinct from all the others in their group in order to minimise collusion. However, because of the high number of students and consequently the considerable effort required in setting and marking, the coursework element had, more often than not, been reduced to one or two in-class tests; indeed to keep staff effort to within reasonable limits even the in-class tests would be of such a form that scripts could be optically read (in other words the coursework had been reduced to a tick/cross exercise). The advantage of this is that students could have almost instant feedback and they would know how well they scored on each part of the test. Also, staff could easily analyse the results and assess common failings allowing time for remedial work before the end-of-semester examination. However, this is far from an ideal situation, especially from the student perspective.

assessment strategies should be at the heart of learning and teaching strategies

In order to improve their (the students') learning experience we felt a different approach was required. Obviously if staff were to keep their administrative effort under control, different assessment strategies and modern technology would need to be employed.

Each year, Loughborough University admits in excess of 650 engineering students following essentially the same mathematics syllabus. The commonality is particularly marked at first year level, in which much time is spent reviewing A-level mathematics and in bringing all our students up to the same knowledge-base level. The students are from six main disciplines within the engineering faculty: aeronautical, automotive, manufacturing, mechanical, civil and electrical/electronic students (these broad groupings contain numerous subsets of students following related specialisms). There is therefore considerable scope for efficiency gains especially in assessment. However, our aim is not simply to present mathematics to this group in the most efficient way but to ensure that the learning experience of the typical student is enhanced as well. It should also be said that a major focus is in ensuring that all of our engineers have, at their fingertips, a bank of mathematical methods and techniques they can use in other engineering modules and in their future careers. Inevitably, in this area of work, much of their effort is directed at learning by much repetition of basic material. We do not worry about testing higher levels of understanding of mathematics; it would be unfair to do this as, in terms of mathematical knowledge, our engineering students are only *surface feeders* and experience just three time-tabled hours of mathematics each week.

our aim is not simply to present mathematics to this group in the most efficient way but to ensure that the learning experience of the typical student is enhanced as well

It is instructive to make a rough calculation of the efficiency gains that might be possible. Let us agree that a single piece of coursework is set for each of the 650 students. If the work is to be simply marked with no discursive feedback then perhaps allocating ten minutes per paper is not over-generous. (In mathematics it is possible to mark scripts quite quickly, especially if a tick/cross approach is used, and certainly if the same coursework is set for all.) To mark this single piece of work would take about 110 hours (not counting tea breaks)! If we are to provide an effective learning experience in which students are properly focussed on mathematics throughout a twelve week period we would want to set five pieces of coursework (roughly one for each new topic). This would take approximately fifteen man-weeks of effort on the part of staff - an administrative burden which is clearly unacceptable. Even if the time to mark a single piece of work were cut by two thirds to about 3 minutes per script there would still be five man-weeks of effort required. In practice no such approach would be adopted.

Computer Aided Assessment

At Loughborough we make extensive use of Computer Aided Assessment (CAA) which easily allows students to be tested five times per semester without increasing (their) anxiety and without the consequent increase in staff administration. We employ CAA to *drive* the students' learning. However, even with modern technology not

everything is practical. For example, we have to abandon the desire to supervise tests as the University does not have (nor is ever likely to have) a computer lab stocked with 650 machines - a more imaginative approach is required. We are aware that many students (especially engineers) have their own computers and most will have internet access (all of Loughborough University's accommodation is networked and, for a modest weekly charge, students have fast access to the world wide web). We should exploit this resource.

The testing regime is a significant part of a larger learning regime

A further advantage enjoyed by Loughborough University students resulted from the early adoption of Question Mark for Windows testing package. This package allowed us to experiment over a three year period with question types and modes of presentation. Of course that package did not deliver tests over the web. However, with the adoption of Question Mark Perception at Loughborough we have been able to deliver tests, to large numbers of students, over the web since October 2000. This has proved extremely effective. Our Learning Teaching and Development (LTD) Centre has been at the forefront in using this technology ensuring that only bona fide Loughborough University students (and relevant staff) can access the tests (access is based on staff/ student ID numbers and their passwords).

Testing Regime: Mathematics for Engineers

The testing regime is a significant part of a larger learning regime: the Open Learning project, now superseded by the HELM project. (Details of the HELM project: Helping Engineers Learn Mathematics are reported elsewhere. Interested readers should go to the website http://helm.lboro.ac.uk. Here also screen shots of exam questions are available). During the semester, students are given workbooks for each new topic presented in lectures. These workbooks have been written for self-study so students need not attend lectures or tutorials if they feel they know the topic under discussion sufficiently well already. (Though this option is available I regularly get in excess of 70% of students attending lectures). Most topics take two weeks to cover. The week following is called a test week (during test weeks lectures and tutorials run as normal). The member of staff responsible for testing engineers constructs two tests: a trial test and a coursework test. Both tests have the same number of questions covering aspects of the topic just covered in lectures. Both tests have an identical form, selecting questions randomly from previously created question banks. Although the questions presented in trial and coursework tests may be identical this would be a very rare occurrence. The likelihood is that two students sitting next to each other will see completely different tests (though the subject matter and the level of each question will be the same). Also, the trial tests have feedback pages attached. If a student gets a question wrong on the trial test a single page of feedback is available. The feedback may be generic (addressing the solution of similar problems to the one presented) or specific (in which the solution to the given problem is detailed). The only feedback available on taking a coursework test is the overall score and information on which questions were answered correctly.

A test week is organised as follows:

From 9am on Monday until 9am on Thursday a trial test is available on the web. This test can be trialled as often as the student desires. No record is kept by staff on student performance on these trial tests. These tests, being web delivered, can be taken anywhere in the world. The *reporter* part of Perception has information on the time of access, duration, user input etc. We find some students simply access the trial test to get information on question types and level of difficulty without attempting any questions. Most will make a serious attempt at the trial test at least once, many will seriously attempt the trial test up to five times. Many of them will work in small groups sorting out difficulties with the trial test. A good number seek help with the trial test from staff in our Learning Support Centre. Others will access the test, input spurious answers just in order to get the feedback. We find 90/95% of students engage in some way. We view this as a valuable learning mechanism. The important point to note is that students engage with the learning process at a reasonable level *throughout the semester.*

students engage with the learning process at a reasonable level throughout the semester

From 9am on Thursday until 6pm on Friday a coursework test is available. Students may access this test anywhere, and at anytime within this period. However, they are only allowed to take this test *once*. (At least they are instructed to take the test once. We do not (even though, technically, we could remove the test for an individual student after a first submission) impose this rigidly as a (very) small number will have problems with computers or access to the web and will need to take the test a second time. Most students will only do the test once as instructed. Allowing the test to be available to be taken more than once reduces any extra administration to a minimum).

Although there are many possible question types we generally use just three: numeric input, multiple choice and hot-spot.

Numeric Input: In numeric input (by far the majority of our questions are of this type) we expect students to carry out a calculation, working to a certain precision, and then to input an answer correct to 2 (or sometimes 3) decimal places. We are careful to ensure that questions are so posed so as not to require extensive numerical computation before the answer is obtained. Students complain (and rightly so) if they fail to gain marks for a question they mostly answered correctly but failed at the last hurdle to do the last piece of computation or to round correctly or even to input the number correctly. Typically we would require no more than half to a full page of straightforward working per numeric input question. It is

important to note that students have the facility to review their answers and if necessary to change them before submitting.

Multiple Choice: Multiple choice questions are mainly used to give variety and so as to be better able to give method marks. Generally we do not employ negative marking, however, we do use a "none of these" response as well as four possible answers.

Hot-spot: Hot-spot questions usually take the form of a question together with its worked solution presented on screen. The solution may have a number (including zero) of deliberate errors. The student is expected to use the cursor to point to the first error on the screen. If no errors are found they point to a no errors box. Students generally find this type of question quite challenging.

The mark achieved in the coursework test is recorded and contributes to the overall module mark. Typically each test is worth 6% of the module mark, giving a total of 30% for the coursework element of the module. We find that this level of reward is just sufficient to attract the student to spend time on the assessment.

students particularly like the flexibility this method of assessment offers

What is the effect of this testing regime? Firstly, following extensive feedback exercises we find this method of testing generally popular with staff and students alike. Students particularly like the flexibility this method of assessment offers. They like the facility to trial tests and the possibility of doing tests (to a large extent) when *they* are ready. A recurring complaint about this approach is its *unforgiving* nature: if the final step is incorrect then no marks are awarded even though every other step was accomplished correctly. Our response to that is to point out that CAA questions are relatively straightforward and students would expect to get them right each time. It also encourages careful working which is no bad thing. Staff have very little work to do, except for the administrative work of chasing up students who are not performing well and in collating marks together.

Because of the success of this approach it is tempting to have *all* our assessment in this form. It raises questions about the *need* for a traditional end-of-semester written paper if it can be demonstrated that students are working steadily throughout the semester and acquiring the necessary mathematical skills and techniques.

Delivering tests over the web is convenient but it is not essential. All of our tests (that is, all the questions and all the feedback) together with all the course workbooks easily fit onto a CD. Students could then operate in *distance learning mode* doing the required work and completing tests on the CD. When they are satisfied with their progress they could then submit (via the CD) their test results through the web back to the lecturer. This way students would not need to be *live* on the web for long periods; only the few seconds it takes to upload test results. Though we don't use this approach at the University we do run a distance learning refresher module employing this regime for engineers who are spending their gap year on work placement.

It would be impossible to deliver the number of tests used at Loughborough without the use of CAA. Of course there is a heavy price to pay in setting up this regime. The institution has to be committed in modernising its assessment strategy and in funding the costs of new technology and of software. At departmental level dedicated staff are required who can see the long tern benefits of CAA and who are prepared to put the extra effort in to construct good quality and extensive banks of questions. (In the Summer of 2005, the HELM project will make available, free of charge, all of its learning materials and all of its question banks to all HEIs in England and Northern Ireland)

In the learning regime used at Loughborough we have found the use of CAA can be a most valuable mechanism which drives student learning.

References

Beevers C, Foster M and McGuire G (1989) "Integrating Formative Evaluation into a Learner Centred Revision Course", British J of Educ Tech, **20**, pp 115 - 119.

Beevers C, McGuire G, Stirling G and Wild D (1995) "Mathematical Ability Assessed by Computer", J Comp Educ, **25**, pp 123-132.

Beevers C, Maciocia A, Prince A and Scott T (1996) "Pooling Mathematical Resources", Active Learning **5**, pp 41 - 42.

Beevers C, Youngson M, McGuire G, Wild D and Fiddes D (1999) "Issues of Partial Credit in Mathematical Assessment by Computer", Alt-J, **7**, pp 26 - 32

Beevers C and Paterson J (2002) "Assessment in Mathematics" in Kahn P and Kyle J (editors) "*Effective Learning and Teaching in Mathematics and its Applications*", London: Kogan Page pp 49 - 61.

Beevers C (2003) http://ltsn.mathstore.gla.ac.uk/index.asp?cat=137 (accessed 30 August 2003).

Fiddes D, Korabinski A, McGuire G, Youngson M and McMillan D (2002) "Are Examination Results affected by the mode of delivery?" Alt-J, **10**, pp 60 - 69.

McGuire G, Youngson M, Korabinski A and McMillan D (2002) Partial Credit in Mathematics Exams: A Comparison between Traditional and CAA, Proc Sixth International CAA Conference, Loughborough University pp 223 - 230.

Nelson C (1994) "Critical Thinking and Collaborative Learning", in Bosworth K and Hamilton S (editors) *Collaborative learning and college teaching*, San Francisco: Jossey Bass.

Rust C (2001) A briefing on assessment of large groups, LTSN Generic Centre Assessment Series No 12, York:LTSN.

Assessment in MSOR:

Briefing Number 3: Key Skills

Introduction

The Key Skills (variously otherwise called transferable, common or "soft" skills) agenda receives a variable welcome in Higher Education, and there is just as much variety of view in MSOR as in other discipline areas.

In this briefing we explore this issue in general and how it impinges on our own discipline area of MSOR. The briefing will draw heavily on the Generic Series, Booklet 5, written by Roger Murphy (Murphy, 2001), quoting, paraphrasing and summarising from that with permission, but also extending what is said there to relate to the context of MSOR.

We shall begin by discussing what key skills are, why they are an issue in HE, and in particular in MSOR. We shall discuss what opportunities there are for addressing them in MSOR, and advocate that any skills work should be embedded into a student's mainstream academic activity rather than treated as a generic "add-on". We consider issues arising in the assessment of key skills. We also look at a number of examples of how the development of key skills is currently addressed.

originally developed in the context of vocational qualifications .... now more widely an issue

What are the key skills and where have they come from?

Key skills, such as being able to communicate, to work in a team, or to solve problems, are of much interest to employers, and it is not surprising that the educational agenda originally developed in the context of vocational qualifications. However they are now more widely an issue, being "also seen as crucial in equipping individuals for lifelong learning and everyday life in this modern society" (Murphy, 2001).

The Dearing report on Higher Education (Dearing, 1997) recommended that opportunities for key skills development should have a place in all degree level programmes. In fact the Dearing report named a surprisingly restricted list of key skills:

Communication skills

Numeracy

The use of IT

Learning how to learn.

The first three of these were well established but the fourth is worthy of mention. Murphy quotes Dearing directly in this context: learning how to learn is included "because of the importance we place on creating a learning society at a time when much specific knowledge will quickly become obsolete. Those leaving higher education will need to understand how to learn and how to manage their own learning and recognise that the process continues throughout life." (Dearing, 1997).

Perhaps the most widely recognised list of key skills is the one now accepted by the Qualifications and Curriculum Authority (QCA), which settled on a list slightly adjusted and expanded beyond that of Dearing:

Communication skills

Numeracy

The use of IT

Problem solving

Working with others

Improving one's own learning and performance

This list clearly gives some indication of what is informing key skills developments post-16 but pre-HE, and more details are available at www.qca.org.uk.

graduates in MSOR develop a special range of problem solving skills because of the logical and analytical approaches which pervade the discipline

From this basis others have specialised or extended the list of identifiable skills, perhaps in some cases to fit with a developing institution-wide policy (for instance at the University of Luton), or in other cases to fit with enhanced perceptions of skills developed by graduates within a particular discipline area. For instance one might argue that graduates in MSOR develop a special range of problem solving skills because of the logical and analytical approaches which pervade the discipline, and thus a list of key or transferable skills for these graduates should reflect this special property. We discuss this further below.

Why in higher education, and why in MSOR?

Apart from any directly beneficial educational reasons in MSOR (which we move on to later in this section), there are two major external reasons why HE academics need to understand and engage with the key skills issue. Murphy (2001) states the first of these simply: "… students will increasingly enter higher education with experience of key skills programmes, assessments and qualifications". Thus universities need to understand this issue from an admissions point of view. The student experience could come from a range of qualifications, but worthy of note is the A or AS

level Key Skills qualification, which can contribute UCAS tariff points, but which not all MSOR courses would accept as a *relevant* contribution.

The second reason comes from the recommendations of the Dearing report, and a subsequent positive response to it from Vice-Chancellors (CVCP, 1998). Following this, "… all staff teaching in HE should now be aware of ways in which opportunities can be provided for students to develop key skills and all programmes should include opportunities for such skills to be assessed and recorded." (Murphy, 2001)

In fact key skills in some form are now given some prominence through the process of course design and review in all areas, including MSOR. All courses must now be defined through programme specifications (QAA, 2003(a)), and while there is room for some variation in these from one institution to another, the skills agenda is one of the items which must be addressed.

More parochially within MSOR, course design and review is now influenced by the national benchmark standard for MSOR (QAA, 2003(b)). It is worth noting at this point that the benchmark statement recognises the diversity of provision in MSOR, and that this is a valuable feature of the discipline area and course provision within it (in fact the benchmarking team did a highly skilful job in capturing and valuing that diversity!):

"It is appropriate to stress yet again the breadth of the MSOR area and the diversity of programme provision within it."

"There are perhaps few programmes that are entirely theory-based or entirely practice-based. Most programmes have elements of both approaches, and there is a complete spectrum of programmes spanning the distance between both extremes. The position of any particular programme within this spectrum will be clear from its aims and objectives. The important point is that all types of programmes exist currently and are, in their different ways, valuable. This is just one aspect of the diversity of provision in this subject area."

(QAA, 2003(b))

the reactions of MSOR groups in different universities to the key skills agenda are diverse

Within this diversity then, it is perhaps not surprising that the reactions of MSOR groups in different universities to the key skills agenda are diverse as well. However there are some paragraphs in the benchmark statement which make definite statements about the place of key skills in MSOR courses, and some of these are included in Appendix 1.

The case for key skills in MSOR, and for embedding them in mainstream studies

The precise place of key skills in HE is still a matter of some controversy, and MSOR is not aloof from this debate. The diversity of response from different institutions and groups as presented by Murphy (2001) in his choice of case studies illustrates this point, and is echoed in the way that MSOR groups are responding. In some ways the debate mirrors and forms part of a wider period of introspection as the rapid expansion of HE over the last few years provokes re-evaluation of the purpose and nature of a university education.

In MSOR, given the wide acceptance of the benchmark statement within the academic community there is clearly some general acceptance of the value of key skills, although perhaps less agreement over how to embed or to assess students' development of them.

Employers have indeed consistently said they value transferable or key skills, and it behoves us as course designers to bear in mind the future needs of mathematics graduates working in industry, whether as a mathematician or in a more general capacity. Two reports (MathSkills, 1999 and SIAM, 1998) provide evidence of the facts. For instance the SIAM Report on Mathematics in Industry (SIAM, 1998) gives body to the broad generalisations about what employers value, by surveying PhD mathematicians working in industry. It suggests that while teamwork, communication skills, and flexibility for instance in problem solving are valued, the mathematicians themselves felt ill prepared for demands in these respects.

while teamwork, communication skills, and flexibility in problem solving are valued, the mathematicians themselves felt ill-prepared for demands in these respects

Recent publications discuss these findings further. Challis and Houston (2000) discuss the issue of embedding key skills development in the MSOR curriculum, as opposed to treating skills as an add-on, or something to be dealt with in a generic way. "It wasn't until students had something really interesting to talk and write about, did they begin to talk and write in a useful manner." (Challis and Houston, 2000). This paper has the distinction of providing a relevant reference to key skills as early as 1549.

In a later article (Challis et al, 2002), the authors discuss what the key skills are, and issues around embedding them, and they give details of two examples where skills development and assessment have been embedded into the MSOR curriculum. They conclude that:

"...the world of work demands graduates and diplomates who have a sound academic background and who possess the interpersonal skills necessary to use their knowledge effectively. ...

"Transferable skills must be taught explicitly as are all other aspects of the course. ...

"Skills must be embedded throughout the programme and their importance constantly stressed. ...

*"*Skills must be assessed just as the academic elements of the course are assessed.*"*

(Challis et al, 2002)

Indeed this last point leads us on to discuss the assessment of key skills – one of the more contentious and difficult points. Murphy (2001) suggests that while many parts of HE actually give value to key skills, they are not always explicit about learning opportunities, assessments and records, nor good at providing evidence that their students develop them. Assessment of these skills is the biggest challenge: "… assessing such skills in a way that is authentic, manageable and meaningful is not a simple exercise."

if you want to convey to students that you value something then you have to assess it

In MSOR we must ask and answer these questions: should key skills be embedded or separate? Should assessment of the so-called transferable skills be integrated into the rest of the course or programme assessment? Challis et al (2002) give a clear response as above. Others may feel that integrating the skills assessment distorts the integrity of their award. However, working on the principle that if you want to convey to students that you value something then you have to assess it, this issue must be addressed. There is no simple answer, and many solutions are currently being tried, and evaluated, for instance those reported in the case studies in Murphy (2001). Perhaps there is no entirely satisfactory answer within our loveable but arcane UK system of awarding degree classifications. A system in which a full student profile is the outcome of a qualification would allow more possibilities, but perhaps the fact that all universities must provide student transcripts of all marks will help.

Key skills assessment

Murphy (2001) makes some generic points about what constitutes good assessment of key skills, and since these are as applicable to MSOR as to any other discipline they are summarised here.

They are challenging to assess because if they are transferable then they must be tested in various contexts to demonstrate this, and indeed particularly in new contexts; qualities acquired over a range of activities need to be assessed in a holistic way; reporting the results must be open and jargon free; students do not value separate exercises on skills development (see discussion of embedding above); and it would be useful to work on and develop the synergies between skills development and the deep learning/reflective practitioner developments (see the Guide for Lecturers).

They are challenging to assess

Good schemes of key skills assessment allow collection of evidence from a wide variety of sources both within and outside the course; link summary assessments of key skills to specific evidence (perhaps through a portfolio – see Briefing number 4 on Portfolios); show development of a skill over time; and inform the learner as well as providing a summary for others.

Poor key skills assessment schemes use the wrong tools such as timed tests; use "tick box" type approaches which make no link to evidence; present evidence but without evaluation and reflection; refer too specifically to the particular culture in which the tests happen; and are inconsistent.

Some final thoughts on key skills in MSOR

We finish with some musings on opportunities for development and assessment of Key Skills in MSOR. The lists are by no means comprehensive, but may provoke ideas. Obvious places where students will have opportunities for skills development include an industrial placement (if one is present), mathematical modelling activity, and project type work. More unexpected opportunities may arise, for example in engineering mathematics modules. An overarching structure for assessment of the skills can be provided by the use of personal logs and a portfolio of evidence as discussed in Briefing number 4 on Portfolios, and also mentioned in the Open University case study in Murphy (2001). More detailed discussions of some of these ideas appear in various references (e.g. Challis et al (2002), Haines and Dunthorne (1996), Kahn and Kyle (2002), Holton (2001)). Let us consider the skills on the QCA list one by one:

Communication: This skill can be developed through exercises incorporating listening, reading, studying words, pictures, and data, communicating the answer to a range of people and not just peers, in suitable form(s) (e.g. report, article, poster, oral presentation, etc.). Vehicles could range from specific exercises such as comprehension tests after reading, through standard mathematical modelling reports, to final year project reports and perhaps oral presentations. In fact the mathematical modelling process provides an ideal vehicle for developing and assessing the full range of communication skills, from understanding the problem, to communicating validated conclusions, perhaps to non-experts.

the mathematical modelling process provides an ideal vehicle for developing and assessing the full range of communication skills

More unexpected opportunities may arise for instance in engineering mathematics, where a frequent complaint is that "students can do the maths in Maths but not in Engineering". One possible integrating activity would be to set a modelling case study, and award marks for both "doing the Maths" and for writing up a report on the case study.

Numeracy: Obviously one would expect all mathematics graduates to be highly numerate, but this raises the question: "what is the difference between numeracy and mathematics?"

Use of IT: The MSOR benchmark talks about "general IT skills, such as word processing, use of the internet and the ability to obtain information". Perhaps one might offer the opportunity to learn to use a spreadsheet or computer algebra system too, and expect an element of transferability here in that one might expect students to be confident enough to learn new packages as and when they need to in the future.

Problem solving: Students learn how to solve problems that have already been set up. They also learn how to set up problems, getting involved in all these stages: formulating the problem, simplifying as necessary, managing oneself through the process of solution solving ("doing the maths"), verifying and validating the solution, and reflecting explicitly afterwards on the whole process. Once more the mathematical modelling cycle covers the whole territory here, via standard modelling exercises as well as larger exercises such as final year projects, and provides an ideal vehicle both for skills development and assessment.

Working with others: In the world of work most problem solving involves teamwork, and while a mathematician will bring special skills to a team, she or he will also have to function as a member of that team and to be aware of the issues involved in that. It is interesting that there is only one very brief mention of group work in the benchmark statement for MSOR. Once more modelling exercises provide an ideal vehicle for developing group working skills.

Improving one's own learning and performance (learning how to learn): This can be the most tenuous of the skills to identify and to assess. Some sort of reflective activity can provide a vehicle for this, one example of which is discussed in Briefing number 4 on Portfolios, particularly concerning the reflective, or PDP, element of the progress file.

Appendix One: Relevant extract from the MSOR benchmark standard statement

This Appendix includes certain statements taken directly from the Benchmark statement in MSOR (QAA 2003(2)), which refer directly to the place of key skills in MSOR provision.

"3.4.1 Graduates from the MSOR area will have acquired many general skills honed by their experiences of studying MSOR subjects. All these subjects are essentially problem-solving disciplines, whether the problems arise within MSOR itself or come from areas of application. Thus the graduates' experiences will be embedded in a general ethos of numeracy and of analytical approaches to problem solving. In addition, an important part of most MSOR programmes is to take theoretical knowledge gained in one area and apply it elsewhere. The field of application is often a significant topic of study in its own right, but the crucial aspect of the process is the cultivation of the general skill of transferring expertise from one context to another.

"3.4.2 A number of general skills are to be expected of all MSOR graduates, though in some cases they are likely to be developed more in graduates from some programmes than others. Even more than in the case of the subject-specific skills, it must be recognised that some are not susceptible to explicit assessment and indeed some are better not assessed so as to avoid creating imbalances.

"3.4.3 MSOR graduates will possess general study skills, particularly including the ability to learn independently using a variety of media which might include books, learned journals, the internet and so on. They will also be able to work independently with patience and persistence, pursuing the solution of a problem to its conclusion. They will have good general skills of time-management and organisation. They will be adaptable, in particular displaying readiness to address new problems from new areas. They will be able to transfer knowledge from one context to another, to assess problems logically and to approach them analytically. They will have highly developed skills of numeracy, including being thoroughly comfortable with numerate concepts and arguments in all stages of work. They will have general IT skills, such as word processing, use of the internet and the ability to obtain information (there may be very rare exceptions to this, such as distance learning students studying abroad in countries where IT facilities are very restricted). They will also have general communication skills, such as the ability to write coherently and communicate results clearly."

References

Challis N V and Houston K (2000) "Embedding key skills in the Mathematics Curriculum", *Capability* 4 (3), pp 7-10.

Challis N V, Gretton H W, Houston K and Neill N (2002) "Developing transferable skills: preparation for employment" in Kahn P and Kyle J (editors) *Effective teaching and learning in mathematics and its applications*, London: Kogan Page, pp 79-91.

Holton D (2001) (editor) *The Teaching and Learning of Mathematics at Undergraduate Level: an ICMI Study*, Dordrecht: Kluwer Academic Publishers.

Kahn P and Kyle J (2002) (editors) *Effective teaching and learning in mathematics and its applications*, London: Kogan Page.

MathSkills (1999) The MathSkills Discipline Network http://www.hull.ac.uk/mathskills (accessed 30 August 2003).

QAA (Quality Assurance Agency for Higher Education) (2003(a)) *Policy on programme specifications* (accessed 29 August 2003) http://www.qaa.ac.uk/crntwork/progspec/prog-spec-contents_textonly.htm

QAA (Quality Assurance Agency for Higher Education) (2003(b)) *Subject benchmark statements: Academic standards - Mathematics, statistics and operational research* (accessed 29 August 2003). http://www.qaa.ac.uk/crntwork/benchmark/phase2/mathematics%5Ftextonly.htm

SIAM (Society for Industrial and Applied Mathematics) (2003) The SIAM Report on Mathematics in Industry, http://www.siam.org/mii/miihome.htm (accessed 30 August 2003).

Assessment in MSOR:

Briefing Number 4: Assessment of Portfolios in MSOR

Portfolios and MSOR

Although one may be aware of the use of portfolios in assessing student work in subject areas such as the fine arts, it may come as a surprise that one of the briefings in this series concerning assessment in MSOR should contain discussion of such a means of assessment.

However we encourage the reader to read on, and to reflect on the points raised here about the wider context, and the possibility that portfolios can make a contribution in addressing some of the issues we face, in assessing certain kinds of work in MSOR, in helping students to reflect on and improve their learning, and in helping students to prepare themselves for employment. We begin by talking generally, and later on give specific examples, not necessarily as proven examples of good practice, but against which to measure one's ideas and thoughts.

This briefing will refer significantly to LTSN Generic Centre booklet "A Briefing on Assessment of Portfolios in Higher Education" by David Baume (Baume, 2001), variously quoting, paraphrasing and summarising from that with permission, but extending what is said there to relate it to the context of our own discipline area, and to other national developments.

In particular we shall: discuss what constitutes the current understanding of good practice in assessment, and how that applies to portfolios; look at a small number of examples of where they are already being used in MSOR; consider briefly where else they may be useful; and finally look at how a portfolio of work can make a contribution to a student's progress file and particularly the PDP element of that file – which all courses in all HE institutions are required to make operational by 2005/6.

a portfolio of work can make a contribution to a student's progress file

What is a portfolio?

Essentially a portfolio is a collection of items rather than a single piece of work. There are certain key features that should be present.

Evidence of work done, either in a particular module (course) or indeed across a student's whole programme of study. This evidence could be a collection of work created and assessed elsewhere, or it could have been created specially for the portfolio. Either way it should be chosen to support the claim for learning that students are making.

A certain minimal level of labelling of this evidence, including identifying those people who created each piece of work and their contribution, its date of production, its title and what it is, and any technology required to access it.

Good structuring and signposting, including at least a contents page, and clear linking if hypertext is used to access a web- or CDROM-based portfolio.

Critical reflection or commentary, probably created specifically for the portfolio. This reflective writing serves the purpose of contextualising and making sense of the evidence. It is where students can make claims about what the evidence says about their learning, and indeed it can provide a self-assessment mechanism to allow them to reflect on this. It is by no means a secondary part of the portfolio and can play a crucial role in enhancing learning.

What is it for?

Filing: a student can use a portfolio to organise their work, in one module or across their whole programme.

Learning: in reviewing and presenting evidence, students will reflect on it and become aware of what they have or have not understood, how much they have learnt, and where they have gaps which they need to remedy.

Assessment: a portfolio may be presented in various ways for assessment. For instance it may be that the individual pieces of work are assessed individually, or perhaps some of the work is assessed in an interim formative way, before review and final submission of the portfolio.

Evidence for employment: a student can present their portfolio to prospective employers, to give life to what they experienced and learned during their qualification.

One important and useful feature of a portfolio is that it gives an opportunity for students to present what *they* value, and to say why they do so.

(Critical reflection) .... can play a crucial role in enhancing learning

Portfolio assessment

Having said a few words about what and why, inevitably one will ask how a portfolio may be assessed. Baume (2001) has a great deal to say about what generic features should be present in any good assessment scheme, and discusses how a portfolio style of assessment can help these features to be present. We summarise this discussion here, adding some asides of our own, before going on to consider examples of some current MSOR use of portfolio assessment. Baume says that any good assessment scheme must have the following properties: validity, reliability, fairness, openness, efficiency in use of both student and tutor time.

Validity

Assessment should test whether a student has achieved the goals of the module, course or programme. In MSOR, while the three-hour unseen written examination is commonplace, there are considerable limitations to what can be achieved and measured in such a format; indeed the same is true of traditional "right or wrong" coursework. There are particular areas where this is an issue, for example in assessing industrial placements and related work, or mathematical modelling activity. Portfolios can make a contribution here. In an extreme case, as Baume (2001) quotes from one of his own papers: "The portfolio can be the student's response to what is probably the most valid assessment task – 'Show that you have attained the learning outcomes of the course.' (Baume, 2000)". We shall discuss this point in more detail later when we look at particular examples in MSOR.

Reliability

Assessors should agree closely on their academic judgements on a piece of work. In traditional mathematical assessment there is a feeling that examiners are highly likely to agree, since a piece of mathematics is either right or wrong, and if it goes wrong, we are used to allocating proportional marks on a commonly understood and agreed basis. Thus reliability has been an issue for more discursive disciplines to worry about but not us. However now that many MSOR courses do include some learning outcomes with a more discursive flavour, perhaps encompassing project, modelling and related work, and the requirement for instance for report writing, then we must face this issue.

Fairness

Baume (2001) has this to say: "Fairness is a profoundly subjective concept, and one that raises strong passions (and from a very early age!) Validity and reliability contribute to perceived fairness. So, often, does the principle that equal marks should reward equal effort." This raises the dilemma of the extent to which the precise nature of a portfolio should be closely specified, as opposed to being within the control of the student. For example maximum overall length could be specified, but further freedom is an issue we consider further below.

Openness

"A good assessment system has no secrets. Students and staff know and understand the course outcomes, the assessment criteria and the assessment process" (Baume 2001). In traditional MSOR assessment this is not a problem as there is an obvious rational and objective basis for marking. However with the use of instruments such as portfolios, it is important to recognise the need for clarity and explicitness in stating the criteria for what constitutes good or not so good work. This is to enhance the perception of fairness in the final grade; it is also to ensure that any interim feedback given on components of the portfolio is genuinely useful in supporting students to improve their work before final portfolio submission.

Efficiency in use of both student and tutor time

Producing a portfolio is very time-consuming for a student, but as Baume (2001) says, "the necessary acts of production, selection, critical judgement and reflection are, I believe, profoundly educational and developmental." As for staff time, assessing a portfolio can be a long job, although this can be worth it if the portfolio is a major part of the overall assessment. To make it more manageable, one can specify an upper size limit. Also it is often the case that some of the material in a portfolio has been previously assessed and the subject of feedback, so in that case a complete reading may not be necessary, but perhaps only a look to ensure that any claims made by a student in their commentary are true.

As for the mechanics of the assessment scheme, there are other key issues identified by Baume (2001) which need to be addressed.

How detailed a judgement?

In MSOR we are used to designing marking schemes for assessments in which marks are broken down into small chunks. Such an approach may not be useful with portfolio assessment, and there can be difficulties in making gradations too fine. Baume identifies three particular sources of difficulty: assessors may not agree on what distinguishes one level from another; there may be more than one way for a particular level of performance to be achieved; assessors might be able to agree a grade, but without being able to articulate why!

This point relates closely to the next one.

How to reach a judgement?

In MSOR, we are primarily used to applying a bottom-up or outcomes based approach to reaching a final grading, using a broken down mark scheme. As Baume points out though, all assessors carry their own standards, and in assessing work with a more discursive element, they will often take a top-down or holistic approach to express their overall feeling about the grade they feel a piece of work deserves. Indeed Baume points out that some assessors iterate between these two approaches:

*"*They check whether the rules-based application of individual judgements on elements of assessment 'adds up to' their overall, holistic, judgement. If these match, then all's well. If not, the assessors may modify their holistic judgement to match the judgement that the bottom-up approach gives; or they may alter individual judgements on elements to match their holistic judgement, or make some combination of these two kinds of change*."*

Whose work?

Some assessors mistrust "coursework" since you cannot know for sure that it is wholly the work of the student (see the briefing on Plagiarism). Baume suggests various ways of addressing this issue. You can explicitly reward appropriate referencing. You can ensure that students must set their submission in the context of their own unique experience and knowledge – for instance where an industrial placement is being assessed. You can require critical

reflection, which by its very nature is individual, and identifiable as such.

The particular issues around plagiarism in MSOR – for instance that a piece of mathematical work is either right or wrong, do not necessarily apply here, and as Baume points out: "Plagiarism becomes progressively more difficult as the task becomes more particular to the individual student, to their own interests and experiences – another argument for assessment via portfolio."

How should students and assessors be briefed?

Both students and assessors are likely to be unfamiliar with a form of assessment such as a portfolio. (Indeed the same is true of other discursive work such as projects, reports, and learning diaries, where students have been known to vent their feelings by saying "I came here to do mathematics, not write essays"!) Therefore perhaps more work than usual has to go into explaining the criteria by which a piece of work will be judged, maybe showing examples of former work and using it to generate discussion of its merits and what it demonstrates. Also since assessment is a central process which generates strong emotions – fear, indignation, joy, anxiety – one must also be prepared to have to argue the case for an unfamiliar assessment mechanism which may at first seem threatening.

What happens if one fails?

A designer must address this question. If the portfolio is designed around a requirement to satisfy a number of discrete outcomes, then a student may be allowed to resubmit just those parts of the portfolio relating to any outcomes not achieved. If the portfolio just requires an overall pass, then it may be appropriate to ask for resubmission of the whole portfolio, perhaps with specified improvements.

Examples

Professional Practice

If a course contains a professional practice element such as a work placement, then a portfolio style of assessment of that practice can be helpful.

Professional Studies

At Sheffield Hallam University, those students who do not do a work placement have to take a module called Professional Studies. This aims to help them to consolidate their professional skills, and to prepare for employment by studying mathematical issues in industry. Some details of the learning outcomes and assessment specification are included in Appendix 1A. The module is assessed by portfolio.

Initially there was a fairly free and generic portfolio specification, but students found this hard to deal with, and it has been found necessary to be more specific about the tasks they are asked to perform and the evidence they should present in their final portfolio. Now four components are required:

An individual mathematical article, to encourage communication skills.

Individual work around the issues of job applications.

Group work with the aim of investigating what mathematicians actually do and the skills they require.

An individual review and action plan, to encourage reflection on strengths and weaknesses, and how to deal with those weaknesses.

These tasks provide consolidation of and focus for skills work; awareness of employment issues for mathematics graduates; and a chance to reflect on one's own position. Two detailed points are worthy of mention. Within the constraints, it is still possible for students to express themselves by selecting what work they present to give evidence of achieving the learning outcomes; and with some items such as the mathematical article, there is interim feedback available to help form the final version of the portfolio.

focus for skills work; awareness of employment issues for mathematics graduates; and a chance to reflect on one's own position

Modelling

Modules in Mathematical Modelling are perhaps often effectively assessed through the portfolio of work a student presents. Although there is often no choice in what to present, work is often presented sequentially as it is done, and there is no overall reflective element. However there may be room to consider, and educational value in, introducing an element of student input in choosing what work will best illustrate that they have achieved the learning outcomes, and allowing them to say why they think it does so. In a more radical form, perhaps there may be value in allowing students to claim that they have satisfied learning outcomes which were not even there in the original module specification, although this would be an unfamiliar situation in MSOR!

Wider possibilities

The idea of a portfolio can make a contribution to how students view, integrate and develop their work, and their working habits, across a whole course or programme, and so can contribute to the development of Progress Files and personal development planning (PDP), and this is discussed in more detail in the following section.

Portfolios and progress files

At the time of writing there are important national developments which will impinge on all discipline areas including MSOR, concerning the development of progress files for all students on all HE courses, and to which we suggest the notion of a portfolio can make a useful contribution.

We might first ask what the national view is of what constitutes a progress file. The QAA website (QAA, 2003) has this to say:

(A Progress File is …)

"A transcript – provided by the institution; an individual's personal records of learning and achievement; a means by which students can monitor, build and reflect upon their personal development; the term "personal development planning (PDP) is used to denote this process and HEIs are expected to have their own policies in place by 2005/6."

Note that there are three components here: the transcript, which is a formal statement of marks and is not considered here; a "personal record of learning and achievement", which could be identified with the evidence base of a portfolio; and a "means by which students can monitor, build and reflect upon their personal development". This last, PDP, element can be identified with the critical reflective element of a portfolio.

The LTSN Generic Centre website (LTSN Generic Centre (2003(a))) emphasises the same points:

"The Guidelines for HE Progress files ...., suggest that the PDP element of the policy objectives should be operational across the whole HE sector and for all HE awards by 2005/6."

Two useful articles to be found on this website are Guide for Busy Academics No 1: PDP, and Guide for Busy Academics No 3: Using PDP to help students gain employment.

There is some recent evidence, with a preliminary report available on the LTSN website, (LTSN Generic Centre (2003(b))) that "Students can improve through the use of learning logs, journals, diaries and electronic portfolios." (quote taken from that site). However there is no evidence emerging that the process helps their employment prospects. At the time of writing, the publication of the full report is awaited.

Having seen that the progress file issue is one that all universities in all courses must address, let us study the key features of the personal record element, and the so-called PDP element, and see how the idea of a portfolio has a contribution to make.

Digging more deeply into the guidelines for progress files, there is significant emphasis on the fact that this kind of activity should be fully integrated with a student's mainstream academic pursuits; it should be linked to the learning objectives and programme outcomes of the student's main academic studies; and it should be implemented in such a way as to gain the support from both academic staff and institution. If this is to be useful then, the subject area academics must become involved in designing any progress file scheme, and the portfolio ideas in this briefing provide one possible basis for such an implementation.

Of course there will be choices to make. One such choice is whether to assess or not. We have said elsewhere in these guides that one strong and pragmatic view is that what you assess is what you get, but one may feel otherwise with reason and, even if one does agree, there is of course more than one way of assessing.

By way of illustrating these rather theoretical ideas, we present an example in Appendix 1B, of how an online web-based portfolio, with reflective element, has been implemented on the Mathematics degree programme at Sheffield Hallam University. As we have said before this is in its early stages of implementation, with the first students to use it only just reaching their final year at the time of writing, so we present it as work in progress, against which one can perhaps measure one's own developing inspirations and ideas.

Conclusion

In conclusion we hope that this briefing proves thought-provoking; that we have conveyed and interpreted the ideas put forward by David Baume (2001) in a suitable way for our own discipline area, and that we have extended the content of that generic guide to raise the issue of progress files in a constructive and helpful way.

We close by quoting some of the words from Baume's own conclusions:

"I have suggested that portfolios can be used effectively, in conjunction with other forms of student work, to prompt, support, integrate and then assess student learning. Good practice in assessing portfolios has many of the same prerequisites as good practice in any kind of assessment; chiefly, a clear account of what learning is to be assessed, and then a willingness to analyse student work and its assessment and to make informed improvements year on year to assessment practice….

"… Perhaps the best single reason for the use of portfolios is that students value them, as a tangible outcome from and demonstration of their learning."

Appendix 1A Professional Studies at Sheffield Hallam University

LEARNING OUTCOMES:

On completion of this unit the student will be able to:

draw upon the knowledge and experience of others and upon published resources to establish the range of skills and competences demanded in areas of employment relevant to their own future careers;

assess their own skills and competences in relation to these expectations;

demonstrate negotiation and organisational skills through the development of a team activity in collaboration with others, designed to extend and consolidate these skills and competences;

produce a high quality curriculum vitae and letters of application for specific posts;

communicate effectively orally and in writing according to demand of the context.

UNIT ASSESSMENT AND FEEDBACK STRATEGY:

Students will be assessed on their ability to fulfil the anticipated learning outcomes.

Assessment will be based upon a portfolio of evidence of achievement of these learning outcomes structured by specific, and in some cases negotiated tasks which will include a piece of critical writing concerning the role of a mathematician or mathematics in industry, and a presentation.

Students will receive informal tutor feedback throughout the construction of their portfolio. Assessment may include elements of self assessment, peer assessment and tutor assessment.

Appendix 1B The web-based portfolio and learning log at Sheffield Hallam University

On the Mathematics degree programme at Sheffield Hallam University, all students create an archive of their work, and also regularly complete an online, reflective learning log (Waldock, 2002, Challis et al, 2003). This exhibits the main features, the personal record and the PDP element, of a progress file. It also maps onto the idea of a portfolio as considered here.

A brief description of the main points follows.

The portfolio is electronic- and web-based. All students learn some very basic web page creation and maintenance skills, to help them with the mechanics of storing their work. It inspires some to do most of their work on computer; others, if they work on paper, will scan good examples of work into the machine.

The activities and habits formed in creating this become fully integrated into mainstream academic work.

The reflective learning log is completed at least weekly, with a very small contribution to the mark from each entry. Staff feel they may be detecting some evidence of year-on-year increase in sophistication in the comments made in the log.

Students create and regularly update a high quality curriculum vitae, which is useful for placement or graduate job applications, and plays a part in encouraging regular reflection.

All elements bear some credit awarded as part of the key skills element of the degree.

The portfolio is not currently linked to university's formal transcript of student performance. There is a strong feeling that full integration with central university systems is not necessary, and that the integration of the reflective and recording elements into the discipline which the student has come to study provides over-riding value. Students engage with this process because they see it as part of the course, and not as an add-on.

References

Baume D (2001) *Assessment: A Briefing on Assessment of Portfolios,* LTSN Generic Centre Assessment Series, LTSN Generic Centre

Baume D (2000) Dialogues: *Assessment. Educational Developments.* 1(3)

Challis N, Gretton H and Waldock J (2003) "Web-Based Assessment of Student Progress Files" in CAA in Mathematics [online], LTSN Maths, Stats & OR: Birmingham http://ltsn.mathstore.ac.uk/articles/maths-caa-series/dec2003 (accessed 15 December 2003)

Learning and Teaching Support Network Generic Centre (LTSN Generic Centre) (2003(a)) http://www.ltsn.ac.uk/genericcentre/index.asp?id=17577 (accessed 28 August 2003).

Learning and Teaching Support Network Generic Centre (LTSN Generic Centre) (2003(b)) http://www.ltsn.ac.uk/application.asp?app=resources.asp&process=full_record&section=generic&id=231 (accessed 28 August 2003)

Quality Assurance Agency for Higher Education (QAA) (2003) (accessed 28 August 2003) [online] http://www.qaa.ac.uk/crntwork/progfileHE/contents.htm

Waldock J ,Gretton H and Challis N (2002) "Using the Web to Enhance Student Learning", *Proc 2nd International Conference on the Teaching of Mathematics*, University of Crete, Hersonissos, http://www.math.uoc.gr/~ictm2/Proceedings/pap199.pdf (accessed 31 January 2004)

Assessment in MSOR

Briefing Number 5: Plagiarism

Introduction

What is plagiarism?

A typical dictionary (Chambers, 1998) defines plagiarise as "to steal from (the writings or ideas of another)". At least one University defines plagiarism as "the misrepresentation of the work of others as one's own (including ideas, arguments, words, diagrams, images or data)." (The University of Reading, 2002) Other sources give broadly similar definitions but with some variation in the detail. Some sources are more explicit about the dishonesty which is implicit in that word "steal" used by Chambers, others lay greater emphasis on the acknowledgement of sources. This variation is important, as it may give contradictory indications to students about what is acceptable. It may also seriously affect what would be satisfactory evidence of plagiarism.

It is clear to all that if a student is asked to do a piece of work, then simply copying someone else's work is not acceptable. More importantly, perhaps, it is not sensible. If there is a purpose in asking the student to do the work it will be to promote or assess the student's learning, and copying contributes nothing to that purpose.

simply copying someone else's work is not acceptable

At the other extreme a student might think up his or her own ideas relating to the task in hand, sort them out and write the result in his or her own words. This total originality clearly involves no plagiarism but in most cases, particularly in Mathematics, Statistics and Operational Research (MSOR), it is unreasonable to expect it. Any realistic piece of student work is going to use sources, ideas and methods from the lectures for the module concerned or from other sources.

we do not require students to acknowledge the sources of the standard ideas introduced in lectures, or the "common knowledge" of the subject

The issue is one of acknowledging this input appropriately, for we do not require students to acknowledge the sources of the standard ideas introduced in lectures, or the "common knowledge" of the subject.

What are we trying to avoid?

Much of the literature on plagiarism concerns essay-type material, although this applies equally to reports and dissertations, which are more often relevant to MSOR. In such a task students will find their own sources of information and ideas. It is unacceptable for them simply to quote, perhaps with slight changes to words, large amounts of text from these sources without acknowledgement; that is copying. If the whole report consisted of such quotations, each duly acknowledged, but with only very slight introduction and conclusions added by the student, then it would be a poor piece of work but not plagiarism. Is your advice to students clear that this work would be judged as poor? A less extreme case would draw from several sources, possibly harmonising the terminology, and would quote conclusions from the sources, all duly acknowledged. Again not plagiarism, but not good: the student has done something in discovering other people's attempts at the question, but has not gone beyond their efforts. This is something like what Biggs calls "plagiphrasing" (Biggs, 1999, p129). This effort might be worth a Third Class or, if it draws on several sources, Lower Second Class, mark. However, this sort of literature search might be quite highly regarded outside the academic world: the student has found out what others have done without reinventing the wheel. (The authors plead that the cliché is so much in common use that it does not require acknowledgement.) Unless warned about this, a present day student might well be aggrieved by a poor mark for such an effort.

Applying an idea or technique from one area of study to a problem in another area is wholly commendable

More difficult is the case where the "evidence" is quoted, duly acknowledged, from the sources but the argument, although wholly derived from them, is written in the students' words. If the argument were acknowledged it would not be plagiarism; if not acknowledged it is, at least in principle, plagiarism. Inadvertent innocent use of an argument is possible. This is what Evans (2000), quoting Carroll, calls cryptomnesia. However in MSOR we will seldom expect wholly new methods, so some re-use of arguments is expected, but some more stringent plagiarism policies may cast doubt on this.

What are we expecting? Applying an idea or technique from one area of study to a problem in another area is wholly commendable. Here acknowledgement of the source might not only make it more readily understood but emphasise the student's contribution to the argument.

The key to the sort of behaviour that we are trying to avoid in essays, reports and dissertations is clear advice to the student. This is not much different from what we expect in MSOR for a good piece of work: that the student justifies the steps he or she takes, quoting reasons, perhaps theorems

or methods used, in each case. The sources in the essays might be different but the clear attribution of reasons for the steps is not.

Joint or group work

Students may discuss a piece of work they all have to do, something which we ought to commend. They may think through the issues together and come to some joint conclusions. In group work this is explicit and uncontroversial, as the output will be acknowledged as coming from the group. In cases where the students are asked to carry out individual work there is frequently a grey area here.

We cannot, and should not, ban the discussion

We cannot, and should not, ban the discussion. If the discussion considers the issues involved and reaches broad conclusions, if the students acknowledge the joint discussions and then write up their own accounts in their own words reaching their own detailed conclusions, few would cavil. If they all contributed to the discussion and the individual write ups are different, is this not attaining all the outcomes which the assessment was intended to achieve? In essay-type material it is reasonably clear that independent write-ups will be different, but for analysing data or solving a mathematical problem this is less clear: the theory will constrain independent correct work to be very similar, at least for students of moderate ability. If the students' discussions go further and reach detailed conclusions so that, apart from the wording of these conclusions, their submissions are identical in content, is that going too far?

students need to know the bounds of acceptability

If the collusion is openly acknowledged it is not plagiarism, but it might result in a reduced mark; for this reduction to be fair, students need to know the bounds of acceptability.

Problem solving

Copying the answers from another student or from a past set of solutions to these problems is both pointless and dishonest: copying is cheating. The lecturer, of course, is contributing to the cheating in the second case by not bothering to set different problems from a previous occasion. If the lecturer can plagiarise by copying a previous set, why should the students be expected not to follow the example?

Problem-solving of the mathematical or statistical kind raises some issues that need to be made explicit. We have said above that using ideas, arguments, words, diagrams, images or data from another person without acknowledgement is plagiarism. This policy needs to have some common sense injected into it: it would be nonsense if we expected each student to acknowledge the idea of adding every time he or she carried out some addition in the course of a calculation.

We do not need to acknowledge "common knowledge" of the subject. However, given the somewhat draconian stance that is taken in some essay-oriented advice on plagiarism (see for example Pyper, 2003), we in MSOR need to make it clear to the student exactly what is allowable.

We do not need to acknowledge "common knowledge" of the subject

To some extent this meshes with the expectation that students give a reasoned deduction of their conclusions from the given information. This should say why the various conclusions follow from the data: by a theorem or the application of a standard technique, perhaps. This is needed to make the work convincing, but it also makes it clear that originality is not being claimed for the results used (for if so he or she would have to provide a fuller explanation). In such cases the lack of reference to the results is not potentially deceptive, but it may devalue the work. Good work gives sources.

Advice to students

"Students are natural economizers", according to Harris (2002), paragraph 1, who suggests they seek the shortest route through the task in hand. Norton et al (2000), p227, give a daunting list of cheating behaviours used at least occasionally by students. It is up to lecturers to advise them on the best way to achieve the ends they seek and to point out not just that plagiarism is forbidden, and may lead to penalties, but the more positive message that good study methods will make it less likely. The key advice has to be:

Sources must be acknowledged in detail, and be clear what is quoted (in inverted commas), or paraphrased (Bloggs claims that ...).

Students need to be helped by a clear statement of what is the "common knowledge" of the subject, for which they need not supply references. It is important that we supply this guidance, for the tradition in MSOR is not to acknowledge these things, a practice that may differ from that in other subjects to which the same institutional policy applies. Your students need the protection of good advice!

Students also need clear guidance on the extent to which collaboration, duly acknowledged, is or is not acceptable.

Your students need the protection of good advice

Van Bramer (1995, paragraphs 1-3, and Race (2001, p17) have clear advice about references, and Northedge et al (1997, pp251-3) about plagiarism generally.

University policy

Clearly the staff in each institution need to adhere to that institution's policy. This applies to plagiarism as it affects students as much as anything else. Plagiarism, however, can be difficult to prove and any ambiguities in the policy,

or policies which cannot be clearly applied to a particular discipline, make the policy unenforceable. Is your University's policy sensible in the MSOR context? Should you be seeking to have it modified? Have you explained and amplified it to your students?

University policy, however, applies not only to students. Do the lecturers adhere to the plagiarism policy in their lecturing? Do you acknowledge what you have used from others in the way that your students are expected to and lead by example? Do you accept work from students who have not given their reasons justifying a step in an argument? It is not a large step from this to not acknowledging sources.

Plagiarism policy requires clarity about what is and is not acceptable practice

What to do?

Apart from advising students, clearly, on what is and is not acceptable, think about the task you have set. Even if the task was appropriate, and the advice was clear, you may still receive work that appears to be plagiarised. Perhaps you can detect this by the usual signs of poor work: unexplained terms, changes in style, or references to results not previously stated. If the sources are not acknowledged the plagiarism may be detected by typing the suspicious phrases into a search engine (e.g. Google) which reads the content of documents, or using the JISC plagiarism detection service. Both JISC and Harris (2002) have advice on this. At any rate the alleged plagiarism must be investigated following your own institution's policies.

Conclusions

Plagiarism policy requires clarity about what is and is not acceptable practice, and this might usefully indicate what is good practice. It requires clarity on what is the "common knowledge" of the subject, that need not be acknowledged. Above all, it requires that staff and students know about these things. Lecturers can help by setting tasks which reduce temptations to plagiarise, and by being quick to point out that work which appears to be very derivative is poorly regarded.

References

Biggs J (1999) *Teaching for Quality Learning at University*, Buckingham and Philadelphia: SRHE and Open University Press.

Evans J (2000) "The New Plagiarism in Higher Education: From Selection to Reflection", *Interaction*, **4,** http://www.warwick.ac.uk/ETS/interactions/vol4no2/evans.htm. Evans quotes Carroll, RT *Cryptomnesia, The Skeptics Dictionary*, 1998, http://www.dcn.davis.ca.us/~btcarrol/skeptic/cryptomn.html (accessed March 20, 2000)

Harris R (2002) *Anti-Plagiarism Strategies for Research Papers*, http://www.virtualsalt.com/antiplag.html (accessed 6 March 2003).

Northedge A, Thomas J, Lane A, Peasgood A (1997) *The Science Good Study Guide*, Milton Keynes: Open University.

Norton LS, Tilley AJ, Newstead SE and Franklyn-Stokes A (2001) "The Pressures of assessment in Undergraduate Courses and their Effect on Student Behaviours", *Assessment and Evaluation in Higher Education*, **26,** pp 269-284

Pyper HS (2003) "Avoiding plagiarism, advice for Students" http://online.northumbria.ac.uk/faculties/art/information_studies/Imri/JISCPAS/site//pubs_student_avoiding.asp (accessed 6 March 2003).

Race P (2001) *Assessment: A Guide for Students*, York: LTSN Generic Centre.

Stefani L and Carroll J (2001) *A Briefing on Plagiarism*, York: LTSN Generic Centre.

The University of Reading (2002) Examinations Office, *Statement on academic misconduct*, http://www.rdg.ac.uk/Exams/academicmisconduct.htm (accessed 5 February 2004).

Van Bramer SE (1995) *What is Plagiarism?*, Chester, Harrisburg and Wilmington: Widener University. http://science.widener.edu/svb/essay/plagiar.html (accessed 6 March 2003).

Bibilography for further reading

Larkham PJ, "Exploring and Dealing with plagiarism: traditional approaches".
Sound advice, predominantly for essay-like materials.
http://online.northumbria.ac.uk/faculties/art/information_studies/Imri/JISCPAS/site/pubs_goodprac_larkham.asp
(accessed 6 March 2003).

Oxford Brookes University, Good Practice Guide. Good advice and various related papers, including some brief and useful recommendations.
http://online.northumbria.ac.uk/faculties/art/information_studies/Imri/JISCPAS/site//pubs_goodprac_guide.asp
(accessed 6 March 2003).

Park C (2003), "In Other (People's) Words: plagiarism by university students - literature and lessons", *Assessment and Evaluation in Higher Education*, **28,** 472 - 488. A considered and balanced account, with an extensive bibliography.

Pyper HS, "Avoiding plagiarism, advice for Students". Sound, thorough advice on referencing material and avoiding plagiarism. Pyper perhaps goes further than many in his assertion that taking a single phrase or a sentence is plagiarism: many of us will have inadvertently used a good phrase without malicious intent.
http://online.northumbria.ac.uk/faculties/art/information_studies/Imri/JISCPAS/site//pubs_student_avoiding.asp
(accessed 6 March 2003)

Stefani L and Carroll J, *A Briefing on Plagiarism*, LTSN Generic Centre, 2001. A good outline of the issues, largely from the lecturer's point of view, and leads to useful sources.

The Joint Information Systems Committee (JISC) Plagiarism Advisory Service gives much more advice, including a service for detecting plagiarism.
http://online.northumbria.ac.uk/faculties/art/information_studies/Imri/JISCPAS/site/jiscpas.asp
(accessed 6 March 2003). Within this site there is the JISC good practice guide (http://online.northumbria.ac.uk/faculties/art/information_studies/Imri/JISCPAS/site/pubs_goodpracguide.asp), the papers by Larkham and Pyper mentioned below, and recommendations from the *Oxford Brookes Good Practice Guide*.

Van Bramer SE, *What is Plagiarism?*, Widener University, 1995. http://science.widener.edu/svb/essay/plagiar.html
(accessed 6 March 2003.). Simple, down to earth advice about quoting sources.

Assessment in MSOR

Briefing Number 6: Self, Peer and Group Assessment

Introduction

This briefing is based on the Generic Series booklet No 9 – A Briefing on Self, Peer and Group Assessment by Phil Race (Race, 2001).

These three aspects of assessment have much in common, and so are treated in the same chapter. Nevertheless, each is distinct and will be treated separately at first. Self assessment is about developing the ability to assess one's own work; peer assessment is about students assessing the work of other students; group assessment is about both the assessment of the products of group work and the assessment of the process of group working. After a brief discussion of each of these, the briefing looks at a rationale for involving students in assessment and at the advantages and risks involved for the lecturer. It goes on to discuss more fully the assessment of group work, implementing peer assessment, implementing self assessment, and enhancing learning through these strategies.

Self-assessment

Perhaps one of the most important skills that higher education can impart to its students is the skill of self-assessment, the ability to judge one's own work against the standards one is required to work to. This brings with it in later life the ability to "know oneself", to know one's own strengths, weaknesses and limitations. At undergraduate level, the ability to self assess enables the student to carry out project work to a better standard. Of course, like any other skill, it must be taught and nurtured. Race (2001) mentions that self-assessment is particularly valuable when it comes to the assessment of tasks that are highly personal such as reflective logs, diaries, action plans, etc. Normally, in MSOR, if such tasks were set, they would be primarily for formative purposes and would not be awarded a mark or grade. The purpose of the tasks would be to promote deeper learning rather than to rank the students. In MSOR, the principal uses would be in helping students to help themselves to make their project work as good as possible, and in helping the lecturer to gain some insights to the working of a group. And in each of these two uses, self and peer-assessment are closely linked, and both are linked to the assessment of groups.

This brings with it in later life the ability to "know oneself"

Peer-assessment

To state the obvious, peer-assessment is different from self-assessment in that students are making judgements about their fellow students, rather than about themselves. However many of the same skills are required – for example using criteria to assess the standard of work and reading material critically. In addition, though, students learn to give feedback and to give constructive criticism with due regard to the person's feelings. Peer-assessment can be used in many situations in MSOR – the assessment of student presentations and posters, the assessment of written reports, the assessment of group work and the assessment of homeworks and attempts at problem sheets. Sometimes peer-assessment can be used for summative purposes. If it is high stakes assessment then students should be well trained in whatever assessment methods they are using. Usually peer-assessment is used for formative assessment and for the training of students as assessors.

Group assessment

This can apply to the assessment of the *products* of the working together of a group by either the lecturer or student peers, and to the assessment of the group-working *process*, again by the lecturer or by student peers. Assessment of the group-working process is a situation when peer-assessment often gives the best insights.

Rationale for involving students in their own assessment

Race gives seven reasons. Which reasons are the most important or the most relevant depend on the circumstances.

1. Because students are already self-assessing and peer-assessing quite naturally and informally

As Race says, students learn a great deal from one another, both in class and out of class. They compare notes and solutions and ask each other how to do the problems and to criticise their projects. Including self and peer assessment in the teaching and learning process legitimises their informal working together, and can help them do it more efficiently.

2. Because tutor assessment is not sufficiently valid, reliable or transparent

There are situations when students are in the best place to carry out assessment, such as in group work. They are also in the best place to carry out some aspects of self-assessment, the "private" things like reflective diaries. Sometimes, when students have been thinking deeply about something, they are in a good position to offer comments to their peers.

3. To deepen students' learning experiences

This is probably the most important reason for including self- and peer- assessment in MSOR. One of the aims of the module or course is to enable students to "know themselves".

applying criteria to assess their peers' work helps them to see other ways of doing things, and to see ways of not doing things

Requiring students to reflect on the criteria describing "good performance" and to apply these to their own work helps achieve this aim. Furthermore, applying criteria to assess their peers' work helps them to see other ways of doing things, and to see ways of not doing things, from which they can learn.

4. To let students in to the assessment culture

Getting students involved in peer-assessment can give them insights into how lecturers assess them. It helps them see how the lecturer's mind works in assessing evidence and in applying a marking scheme to an examination paper. This all adds to making assessment more transparent.

5. To help students towards becoming autonomous learners

Again, this is to do with enabling students to "know themselves". Self-assessment in particular can play a vital role in getting students to reflect on their own learning and performance. They can then judge how their learning is progressing and they can ask for help if that is required.

6. To help students development skills relating to life-long learning

When a graduate enters the work of work, they will soon be working without close supervision and will be expected to put in a good performance. They will need to know how to self-assess so that they can have confidence in their ability to judge their own work. As they progress in their career, they will be in supervisory positions and will be required to make judgements about those they supervise. Experience in peer assessment will help them do this, even though their employer will probably give them some training in supervising others.

7. To help students gain much more feedback than would otherwise be possible

This is particularly the case when there is a large class and not enough tutorial assistants to help the lecturer. Feedback from peers may not be as authoritative as that from lecturers or tutors, but there is much to be gained from more feedback than the restricted amount they might get from the lecturer.

Advantages and risks for the lecturer

As far as the tutor is concerned, there are some advantages and risks. In some circumstances the tedium of assessing many homeworks may be relieved by peer-assessment. But if this is to be used summatively, it will require student assessors to be trained and their work checked and validated. Students reflecting on their own work can open up productive dialogue between student and tutor. If feedback is given on a student's self-assessment, then, not only are they getting feedback on their performance, but also on their assessment of their performance. This should enable students to be better at both – their doing and their assessment of their doing, and it should give the lecturer some insights into the strengths and weaknesses of a particular cohort. Race includes the observation that most tutors who have involved students in self and peer-assessment report that final student performance is better than expected.

not only are they getting feedback on their performance, but also on their assessment of their performance

When students are inexperienced assessors, their judgements should be used sparingly, if at all, for summative purposes. This is especially true when the tutor's work is under close scrutiny, for example during a QAA inspection. Further problems can arise if the external examiner is not experienced in these modes of assessment. Peer assessment can become the target for "hole-picking". It is wise to choose a sympathetic external, or at least to involve the external fully in discussions when a new mode of assessment is being introduced.

There is also a cost involved in initiating such schemes, and this has to be borne in mind and weighed against the lost opportunities to do something else. But do not under estimate the value of the benefits of these innovations.

Assessing student group work

Some of the benefits of self and peer assessment come when they are used in the assessment of groupwork, so this topic is considered first.

For the assessment of project work, students should be involved in the writing of the assessment criteria. This can be achieved through classroom discussion of the aims of project work, the intended learning outcomes, and the features that go to making a good project. At an early stage, students may not be aware of these features. So the discussion could be led towards the criteria that the lecturer desires, although this must be done in such a way that, at the end of the discussion, students feel "ownership" of the criteria, which would then be couched in language that they would understand.

In MSOR, most project work in the early years is carried out in small groups of three or four. There are two things to assess – the product and the process. The product will usually be a written report or a seminar presentation or a poster. It should demonstrate the group's success in carrying out the investigation and in reporting it. Accordingly assessment criteria for judging the success of the investigation are required, as are assessment criteria for judging communication skills. At this early stage of the students' education, it is probably best to mark a written report as if it had come from one person and, initially, to give that mark to each member of the group. It is possible to ask students to identify which sections of the report each one has written, and to judge them on their respective contributions, but this tends to atomise the report and to lose the factors that make it coherent. And it may be "the luck of the draw" that one of the students has written a section which is particularly easy to write. During a seminar presentation the tutor may assess each student's communication skills, but account has also to be taken of the group's decisions concerning overall content and structure. The publication edited by Haines and Dunthorne, published in 1996 and disseminated widely to

UK Higher Education Institutes, contains examples of assessment criteria for written reports relating to mathematical modelling projects, pure mathematical investigations, statistical problems and other, more general projects. This publication also contains criteria for seminar and poster presentations. These have been tested in use by lecturers and students and have been shown to work well. Assessment criteria for written communication and for mathematical modelling investigations are included in Appendices 1 and 2 respectively. These were first published in Houston et al (1993), which had limited circulation, and appear again in Haines and Dunthorne (1996) and elsewhere.

there is the problem of identifying the rogue students who do not pull their weight in the process

However there is the problem of identifying the rogue students who do not pull their weight in the process and who are carried by the others in the group. It is here that confidential self and peer assessment of the group working process can be used to good effect to moderate the mark assigned to the group in order to give a mark to each individual. Again consult Haines and Dunthorne (1996) for an example of this idea in action. They include the assessment criteria first developed by Goldfinch (1990, 1994) for self and peer assessment of group work, and this is included in Appendix 3. Students are asked to rate their own and their peers' contributions to the work of the group in terms of "Level of attendance", "Ideas suggested" and the other attributes listed in Appendix 3.

The Subject Benchmark Statement (QAA, 2002) suggests that a major project should be part of the final year curriculum for MSOR students but falls short of requiring one in recognition of the difficulties some universities with large classes may have in running a project module. If a project is offered at final year honours level, it should ideally be an individual project, because final year work demands high stakes assessment. When the assessment criteria are being discussed with students, it is useful to suggest to them, that, as their work enters its final stages, they should pair off and peer-assess one other's work, paying particular attention to written presentation.

Implementing student peer-assessment

It is widely recognised (see Griffiths et al, 1994) that peer tutoring, in any of its forms including peer assessment, is a powerful enhancer of learning. Having to teach something to others requires the teacher to have a deeper knowledge than they might require even to pass an examination. They have to think about how to present the material, what difficulties they themselves had when struggling to learn it and how best to explain the material to another struggling learner. Houston (1998) and Houston and Lazenbatt (1999) have experimented with peer teaching through oral seminar presentations, written notes and posters. There were elements of peer assessment in this work in that students used the agreed assessment criteria to judge the oral presentations of their peers. These judgements were fed back to the presenting students solely for formative purposes and benefitted both assessors and those being assessed. Houston (1997) describes a case study wherein the peer judgements of students of the posters presented on project work were analysed and compared with the judgements of the lecturer. The exercise highlights some of the difficulties that could arise if peer assessment is used for summative purposes. Houston concludes, "it was discovered that some students had difficulty applying the criteria in a consistent way, and that their judgements were not consistent with the consensus view. It is recommended that students be involved in developing the criteria themselves, or at least trained in their use, so that they have a greater measure of 'ownership' of them."

peer tutoring, in any of its forms including peer assessment, is a powerful enhancer of learning

Another way in which students might be involved in peer assessment is in the marking of homeworks in class, following the guidance of the lecturer using the board to present a correct solution and allocation of marks. This requires students to read the answers in front of them very carefully to see where the errors occur. Again this should be considered to be a learning exercise and not used for summative purposes without complete scrutiny by the lecturer.

Peer assessment of a "buddy's" final year project should be encouraged before submission. This would help eliminate many typographical and presentational errors by which the author would benefit and the assessor would learn from the experience about their buddy's project and about the sorts of errors that they themselves could so easily make.

Implementing student self-assessment

Race (2001) points out that "while students can self-assess just about any aspect of their work, ... there are some things that are arguably best self assessed, including some which in many respects can only be self-assessed. For example, if students are keeping reflective logs or diaries, an activity which is fairly rare in MSOR, only they know 'how deeply they have reflected while putting together such evidence of their reflection'."

Students should be encouraged always to self-assess everything they do. When doing problems from a textbook, they will almost certainly look up the answer "at the back of the book" to help them. It is more difficult when it comes to assessing written reports and presentations and here they should use the assessment criteria systematically.

Students should be encouraged always to self-assess everything they do

If this self-assessment is submitted with the written material then the lecturer can give feedback both on the material and

on the student's self-assessment of the material, and this can be very powerful. For a presentation it is helpful for students to view a video of their performance, if one can be made. The lecturer with a digicam, a tripod and a remote control can quite easily make a video good enough to help students learn from their performance.

Race (2001) suggests some "prompts for self-assessment dialogues with tutors":

What do you think is a fair score or grade for the work you have handed in?

What was the thing you think you did best in this assignment?

What was the thing you think you did least well in this assignment?

What did you find the hardest part of this assignment?

What was the most important thing you learned in doing this assignment?

Enhancing learning

Race (2001) reports that five factors in particular have been identified which underpin successful learning:

Wanting to learn (intrinsic motivation)

Needing to learn (extrinsic motivation)

Learning by doing (practice, trial and improvement)

Learning through feedback (praise, constructive criticism)

Making sense of what has been learned

Involving students in their own assessment and that of their peers is seen to relate to all of these factors.

References

Goldfinch JM (1990) "Development of a Peer Assessment Technique for Obtaining Individual Marks on a Group Project", *Assessment and Evaluation in Higher Education*, **15**, pp 210-231

Goldfinch JM (1994) "Further Developments in Peer Assessment of Group Projects", *Assessment and Evaluation in Higher Education*, **19**, pp 29-36

Griffiths S, Houston K and Lazenbatt A (1994) *Enhancing Student Learning through Peer Tutoring in Higher Education*, Coleraine: University of Ulster.

Houston SK (1997) "Evaluating Rating Scales for the Assessment of Posters" in Houston SK, Blum W, Huntley I and Neill NT, (editors) *Teaching and Learning Mathematical Modelling*: *Innovation, Investigation, Applications*, Chichester: Albion Publications, pp 135-148.

Houston SK (1998) "Get Students to do the Teaching!" in Galbraith P, Blum W, Booker G and Huntley I (editors) *Mathematical Modelling - Teaching and Assessing in a Technology Rich World*, Chichester: Horwood Publications, pp 45-54.

Houston SK, Haines C and Kitchen A (1994) *Developing Rating Scales for Undergraduate Mathematics Projects*, Coleraine: University of Ulster.

Houston SK and Lazenbatt A (1999) "Peer Tutoring in a Modelling Course", *Innovations in Education and Training International*, **36**, pp 71-79.

Race P (2001) *A Briefing on Self, Peer and Group Assessment*, LTSN Centre Assessment Series No 9, York: LTSN Generic Centre.References

Goldfinch JM (1990) "Development of a Peer Assessment Technique for Obtaining Individual Marks on a Group Project", *Assessment and Evaluation in Higher Education*, **15**, pp 210-231

Goldfinch JM (1994) "Further Developments in Peer Assessment of Group Projects", *Assessment and Evaluation in Higher Education*, **19**, pp 29-36

Griffiths S, Houston K and Lazenbatt A (1994) *Enhancing Student Learning through Peer Tutoring in Higher Education*, Coleraine: University of Ulster.

Houston SK (1997) "Evaluating Rating Scales for the Assessment of Posters" in Houston SK, Blum W, Huntley I and Neill NT, (editors) *Teaching and Learning Mathematical Modelling*: *Innovation, Investigation, Applications*, Chichester: Albion Publications, pp 135-148.

Houston SK (1998) "Get Students to do the Teaching!" in Galbraith P, Blum W, Booker G and Huntley I (editors) *Mathematical Modelling - Teaching and Assessing in a Technology Rich World*, Chichester: Horwood Publications, pp 45-54.

Houston SK, Haines C and Kitchen A (1994) *Developing Rating Scales for Undergraduate Mathematics Projects*, Coleraine: University of Ulster.

Houston SK and Lazenbatt A (1999) "Peer Tutoring in a Modelling Course", *Innovations in Education and Training International*, **36**, pp 71-79.

Race P (2001) *A Briefing on Self, Peer and Group Assessment*, LTSN Centre Assessment Series No 9, York: LTSN Generic Centre.

Appendix 1

Communication Skills (Written)

		High			Low	Not shown	Not applicable
W1	Gives a free standing abstract or summary Includes a statement of the problem, which may need redefinition, states the methodology used and gives specific conclusions. This section is free standing, brief and precedes the report itself.						
W2	Gives an introduction to the report States the problem, gives the background to the problem, sets it in context, explains the strategy.						
W3	Structures the report logically Connects main points logically, gives supporting evidence succinctly and concisely.						
W4	Makes the structure of the report verbally explicit Explains relative importance, emphasises important points (Can see what is there without digging). Makes clear the logical function of constituent parts, uses good and consistent internal referencing (labelling).						
W5	Demonstrates a command of the appropriate written language Uses good spelling and grammar. Uses an appropriate style (technical and linguistic).						
W6	Complements logical structure with visual presentation and layout Draws clear tables and diagrams, spaces report so it is easy to read. Makes it aesthetically pleasing, gives a contents list if needed.						
W7	Makes appropriate use of references and appendices Has a good external referencing system, lists references, uses clearly labelled appendices for secondary data, gives bibliography if appropriate.						
W8	Gives concluding section in the main report Appears near the end of the report, makes a summary of the important points which have been logically derived, conclusions represent an adequate solution to the problem, (does not introduce any new information at this point).						
W9	Gives a well reasoned evaluation This could appear before the end. Considers limitations of the solution, discusses possible extensions and makes recommendations for appropriate action.						

Appendix 2

Communication Skills (Oral)

		High			Low	Not shown	Not applicable
O1	Establishes a rapport with the audience Engages and keeps the interest of the audience, eye contact, enthusiasm, personal qualities etc.						
O2	Makes an effective delivery Uses correct speed, pace, tone, pauses, intonation, rhythm, delivery is well rehearsed and spoken, not read.						
O3	Has a good command of spoken English Uses sentence structure correctly, good vocabulary and grammar, appropriate use of appropriate language.						
O4	Gets over the main point using a good structure Explains well, shows evidence of structure, gives appropriate reference to mathematics, shows understanding.						
O5	Gives a clear explanation of the problem and its outcome Gives a clear explanation of the problem, and an effective presentation for the audience.						
O6	Plans and organises overall delivery well Shows evidence of a plan, organisation (at group level), work by a team, group coheres well.						
O7	Makes appropriate use of visual and other aids Chooses the right medium for communication and uses it well. Aids present when needed. Explains the aids well.						
O8	Provides visual and other aids of a high technical quality Sets out OHP transparencies, posters, etc. well, labelling is good, produces a high quality of appropriate aids.						

NOTES:

Descriptors 1 - 4 are for individuals in a group or for individuals making an individual report.

Descriptors 5 - 8 are group descriptors for the whole group or additional descriptors for individuals making an individual report.

Appendix 3

Peer Assessment Form

PLEASE TRY TO THINK BACK THROUGH THE WHOLE EXERCISE AND TO BE OBJECTIVE IN YOUR ASSESSMENT. INCLUDE THE SEMINAR (AND PRESENTATION) AND THE MODELLING PROJECT (AND POSTER) IN YOUR ASSESSMENT.

YOUR NAME_____ GROUP CODE_____

Please award each group member, including yourself, a mark out of 5 for each of the following factors:

MARKS: 5 = much higher than the rest of the group
4 = higher than the rest of the group
3 = same as rest of group
2 = less than rest of group
1 = much less than rest of group
0 = no contribution in this way.

WRITE SURNAMES OF GROUP MEMBERS IN HERE → FACTORS ↓	S E L F				
Level of attendance at group meetings:					
Level of contribution at group meetings:					
Level of contribution outside group meetings:					
Idea suggesting (e.g. getting started, getting around difficulties, forms of validation):					
Extracting something useful from the ideas (e.g. identifying workable approaches, tying ideas together, getting usable equations):					
Keeping the momentum going (e.g. smoothing group operations, good planning):					
Performing tasks (e.g. writing up, calculating, computing, research):					
Abiding by the contract made at the beginning of the semester.					
OTHER: (Specify here any other useful roles performed by members.) ----------					

Assessment in Mathematics, Statistics and Operational Research

Guide for Heads of Department of Mathematics, Statistics and Operational Research

Introduction

This guide owes much to, and indeed in parts quotes or paraphrases directly and extensively (with permission) from, the LTSN Generic Series, Booklet 2, *Assessment: A Guide for Heads of Department* by Alistair Mutch and George Brown (Mutch and Brown, 2001). The aim here is that the ideas therein are expounded, summarised, expanded, and interpreted in such a way as to bring out their relevance in the context of Mathematics, Statistics and Operational Research (MSOR). It will be apparent that many, although not all, of the issues discussed here are common across all disciplines.

The main issue about assessment as far as heads of department are concerned, will be overseeing the creation and enhancement of a departmental strategy, understanding their crucial role in that process, and considering how the process can enhance quality of assessment. This guide appears in a booklet together with the *Guide for Lecturers in MSOR*, since one essential aspect is that all of those involved in a strategy should share a common understanding and acceptance.

As Mutch and Brown (2001) say,

"Crafting is a subtle art. It involves:

- doing some background research to find the appropriate materials and methods;
- the design of a product that is acceptable to its intended users;
- ensuring that the product fits its purpose - and works;
- considering ways of improving it."

Heads of Department must lead the process, forming an atmosphere in which a viable assessment strategy can be created; they must also manage the process, ensuring implementation, monitoring, review, development and enhancement of the strategy. These roles are expanded upon below.

"Crafting is a subtle art"

The emphasis here, then, is on "departmental" strategy, although we might note that what constitutes an appropriate subject grouping will vary. In many institutions, groupings involving MSOR are subsumed into larger bodies: schools, centres, etc. Also notwithstanding earlier comments, there are some things about the MSOR context which inevitably require special consideration.

Perhaps more important than structural details, it is necessary to be keenly aware of the degree of autonomy one has about assessment, and this will vary according to internal arrangements in a department/school, as well as on institutional culture, which can vary from full departmental autonomy, to a managerial culture with a centrally driven agenda, and passing through all points in between. Related to this is having a clear view of one's own attitude to assessment, and an awareness of how that can affect the process.

In what follows, we offer suggestions and prompting questions which may help a head of department in managing and leading the creation, maintenance and enhancement of a strategy.

Before doing this, we discuss what a strategy is, and why it is important to have one at departmental or subject level.

What is an assessment strategy?

A strategy is a set of procedures for achieving a goal or purpose. It usually involves a statement of purpose, guiding principles to achieve that purpose, and a set of procedures based on the principles.

Purposes, principles and procedures

The purposes of assessment are to help students to improve their learning, to provide certification, and to contribute to quality assurance. Further possible subdivisions appear in Figure 1.

It might be noted that in any gentle caricature of traditional approaches to assessment in MSOR, the second of these categories would tend to dominate possibly to the detriment of the first. In contrast to this we should mention that there has also been a high level of innovation in mathematics assessment in recent years, with reports contained in collections such as Haines and Dunthorne (1996), Holton (2001), Kahn and Kyle (2002), or the reports from working groups at the Undergraduate Mathematics Teaching Conference (http://www.umtc.ac.uk). Some more details appear elsewhere in this booklet.

Ideally, identification of purposes leads naturally to identification of principles, which would then suggest appropriate procedures. For example one might select the purpose of grading or ranking a student. One would then expect a statement of a guiding principle concerning say fairness and equity in assessing learning outcomes. The procedures would outline the methods for ensuring this in marking, the use of criteria, interactions with students, and the variety of assessment styles used.

Levels of strategies

Strategies for assessment exist at a range of levels. At the highest level, there is the QAA Code of Practice (QAA, 2002); below that, every institution will have an institutional strategy for teaching, learning and assessment, which will

Figure 1 Purposes of assessment (reproduced from Mutch and Brown (2001)

Learning	To provide feedback to improve learning To motivate students To diagnose a student's strengths and weaknesses To help students develop skills of self-assessment To provide a profile of what a student has learnt
Certification	To pass or fail a student To grade or rank a student To licence to proceed To licence to practice To select for future courses To predict success in future courses To select for future employment To predict success in employment
Quality Assurance	To provide feedback to lecturers on student learning To improve teaching To evaluate a course's strengths and weaknesses To assess the extent to which a programme has achieved its aims To judge the effectiveness of the learning environment To ensure the course is credit worthy to other institutions and employers To monitor standards over time

effectively require that each "department" within the institution must set that strategy within its subject context. This is where the role of the Head of MSOR comes in. Below this, the high level strategies may be interpreted into programme or course strategies, module strategies, and even strategies for particular activities.

Crafting the strategy at departmental or subject level raises various questions to be addressed. Heads of Department must be aware of their institutional strategy, and of how much leeway they have in setting it in context for MSOR. One important question which arises for the Head, when discussing how to interpret a generic strategy, is the extent to which there are special issues in MSOR, which do not apply universally. In some quarters it may be suggested that every discipline thinks it is special, and so there is no excuse for not toeing the line; however there are genuine matters to be dealt with here – for instance many feel that aspects of the QAA generic level descriptors must be interpreted with some care in an MSOR context, and that the MSOR benchmark standard statement is most apposite.

there are special issues in MSOR

Perhaps this is a suitable place to introduce a specific example. First here is an extract from an institutional strategy document (from Sheffield Hallam University):

".... the assessment of student learning outcomes is based on clear criteria, provides appropriate and timely feedback, is both balanced and coherent, and is effective and efficient within the context of different modes of learning."

At the next (school) level this is expanded upon:

"... Students will be encouraged to become confident and independent learners through a variety of support mechanisms which encourage a deeper approach to learning, more effective use of the available staff time, appropriate and rapid feedback, reflection and formative assessment...

... learning outcomes and assessment criteria, in particular pass criteria, will be made explicit to students in order to enhance the effectiveness of learning programmes, and to encourage students to realise their potential."

Finally it reaches the level of mathematics programme document where the ideas are interpreted to lend them meaning within the discipline:

"Research evidence shows that students who are overloaded with assessment tend to operate strategically and do not develop the ethos which we seek to cultivate. At the same time mathematics is a subject which cannot be left until the last minute, and regular engagement with the subject is needed if good progress is to be made. Modules will therefore require students to complete work which will be formatively assessed either formally or informally. .... Written or oral feedback, with spreading good practice in the use of feedback grids, is provided on all continuous assessments.

"Research evidence also points to the fact that outcomes of different modes of assessment advantage different groups within any cohort. In order to minimise the recurrent disadvantage accrued by repetitive use of one mode of assessment, summative assessment will take a variety of forms. Such variety also reflects the need to match assessment mode to the nature of learning outcomes to be demonstrated."

Finally, the grand statements reach their pragmatic realisation through individual module documents. Similar statements occur in other examples; in some cases, more specific conditions are laid down - for instance specifying the proportion of assessment which must be unseen examination and so on. Many institutions or groupings publish their strategies on websites.

Further issues arise from the nature of MSOR as a natural collaborator with other disciplines. Some MSOR groups have retained significant service teaching, and there may be a clash of academic cultures between the MSOR group and the programme/course being serviced. In this case the matter of academic "ownership" of the module can surface: in the case of a disagreement, does the subject group or the course/programme prevail? The same issue can arise in modular schemes, and different cultures can also clash in joint honours. In each case the needs of the subject must be weighed against the need for a coherent student experience. In some cases the issue is dealt with by standardisation of assessment across all modules regardless of their content and desired learning outcomes. This is bureaucratically neat and tidy, but perhaps is not ideal in allowing assessment to fulfil all its purposes.

In any event, there arises the question of how to relate to higher level strategies, and indeed how much freedom to allow in assessment. One can tackle this explicitly or implicitly in the strategy. Mutch and Brown (2001) comment on the effect of different institutional cultures: "the command-control culture may lead to greater consistency within and between institutions, but it can also lead to resistance to change, subterfuge and stifling of innovation." We might add that it can also lead to resistance to innovation which is centrally driven and not owned by academics!

Perhaps the most important question to ask is what role Heads can play in the midst of all these conflicting forces. In fact they can play a crucial positive role, in acting as a buffer or mediator between institution and discipline. They can aim to adapt the institutional strategy to meet the particular needs of MSOR, its staff and students. At the same time they can enthuse staff, and create space and encouragement for reflection, for innovation, and for experiment, by setting the tone of their leadership to allow the possibility of failure followed by learning and improvement, without threat.

enthuse staff, and create space and encouragement for reflection, for innovation and for experiment

Crucially related here are the issues of sharing and dissemination of the results of innovation both good and bad, and associated staff development which may become necessary as a result. Indeed a Head can create opportunities for this, and must be aware of whatever are the current sums of money available to support it. The Head will of course never forget that academics are primarily MSOR specialists who teach, and pass on their enthusiasm for, their subject, and this perspective must inform staff development. Generic staff development is not as productive as that which is subject related.

Why does a department need an assessment strategy?

Mutch and Brown (2001) write: "The short answer to this question is so that a department knows what it is doing with regard to assessment and why it is doing it in that way. A more extended answer would include how a department discovers whether it is doing what it claims to be doing and what mechanisms it has for improving what it is doing."

This raises the point that, in the strange world of quality, quality *enhancement* is becoming more important than simply maintaining the paper trail. Can this be the thought which gives meaning to the quality agenda?

Mutch and Brown (2001) also suggest that, while some might say they will only create a strategy because they are told to, there are other possible good reasons, paraphrased below:

It provides opportunities to reflect upon departmental approaches to assessment;

It deepens understanding and commitment of staff to the processes of assessment;

It helps staff and students improve their approaches to assessment;

It provides support for the exploration of alternative approaches to assessment;

It ensures a degree of consistency in departmental assessment procedures;

It establishes the department's priorities of assessment;

It interprets the institutional assessment strategy within the subject context.

Any strategy must also take the external context into account: accountability and transparency, benchmarking, standards and programme specifications, shrinking resources, more students, wider access and the need to retain and allow achievement, plagiarism, equal opportunities and disabilities, student choice and coherent learning experiences.

It is also worth saying that assessment cannot be separated from learning, curriculum and teaching style, and is a key part of what students react to. What you assess is what you get.

What approaches to strategy should one take?

The choice of approach to take depends on the situation of MSOR within an institution, and whether one wants to work *within* or *on* the existing culture. Possible approaches may be represented on a spectrum.

"Reviews of studies of innovation and change suggest that the more remote the source of directives or of quality demands are, the less likely they are to induce commitment and compliance. Higher rates of successful innovation are more likely to occur in collaborative learning environments than in command-control cultures." Mutch and Brown (2001)

Control ◄——————————————————————► **Freedom**				
Copy the institutional strategy	Strong directive from head of department (or small senior group)	Target-based strategy, with *agreed* targets produced by department.	Guidance-based strategy, with agreed light structural framework.	Abstentionist strategy, leaving all decisions to another lower level
Easy to produce	May work if it includes strong enough targets....	Likely to gain commitment and compliance; relatively easy to manage	Gives freedom; requires mutual trust and professional responsibility, easy to produce and manage, and can lead to innovation	Control is minimal and freedom is maximal ….
Difficult to manage, takes no account of MSOR, unacceptable to staff	....but may lead to superficial compliance and little commitment	More difficult to produce.	Runs risk of being unclear, and may upset central management of university	.... until a crisis occurs, when often things revert to strong directive.

Preparing to develop (or review) a strategy

Choose the people

The Head can (and must) set the tone here. Is this just a necessary chore to be done with minimal effort? Or should it encompass a real process of academic reflection and critical consideration of one of the key processes in the relationship with students?

Possible approaches involve (again on a spectrum):

Write it yourself	Use an existing committee	Persuade a department member to do it	Create a working group
Easy to write but hard to gain acceptance	Again runs risk of non-acceptance	Depends upon the person, how they operate, and the atmosphere in the department	Choice of people is crucial - must be able to work together. Needs a clear brief. Task will take longer but strategy is more likely to gain acceptance.

Do the research

Start with existing departmental documents (or the previous strategy if it is a review). Look at what other departments in your own university do; look also at MSOR departments at other universities. Section 6 of the QAA Code of Practice raises general issues about assessment (QAA, 2000). The Maths, Stats & OR Network provides much support, and so do the other sections of this booklet.

There are also works in the literature from practitioners in MSOR, often the distillation of team work and reported in edited collections (e.g. Haines and Dunthorne, 1996, Houston, 2001, Beevers and Patterson, 2002). Research

should explore what works, and also what reasoning lies behind any method; for instance the current limitations of computer based multiple choice tests must be balanced against their economic attractiveness.

Consult

Consulting with staff may consist of an audit of assessment methods and procedures, informal discussions, open-ended questionnaires, seminars, awaydays. You may receive comments on assessment overload (of both students and staff). These are highly relevant to any viable strategy, and it is important to listen to, and address directly the *real* issues facing those who will have to live with the strategy.

it is important to listen to, and address directly the real issues facing those who will have to live with the strategy.

Write a draft discussion paper

A draft discussion paper can be very useful. Appendix 1 contains a checklist of prompting questions which may help the group producing the draft strategy. This is reproduced directly from Mutch and Brown (2001). It is important that this takes into account and addresses particular features and issues in your own department: for instance, high proportion of mature students; high failure rates; etc.

Redraft

Ideally everyone in the department, and perhaps some elsewhere in the university, should then have the chance to comment on and affect the draft. However it is possible to lose key features of the draft at this stage as comments are incorporated, and it is necessary to guard against this.

The final version and its implementation

The final document should be published, perhaps on your website and/or on paper, and thus its final form must take the audience into account. This audience may include not only students, but also staff who are not specialists in education per se, but in education in MSOR. Thus, make the document user friendly; make it meaningful within the disciplines of MSOR, and avoid specialist educational jargon, which can exclude and repel in equal measure. The implementation process involves adjusting programme, course, or module handbooks, and thus provides a useful mechanism for familiarisation with and embedding of the strategy.

Implementation requires time for people to think and adapt.

Implementation requires time for people to think and adapt. However if the process is managed collectively then it becomes difficult to distinguish implementation from the creation and consultation phases. Staff development is fed and inspired by the whole process as well as by the final product.

Monitoring and evaluating the assessment strategy

Monitoring and evaluating the strategy is an important management task. The Head of Department plays a key role here in ensuring that this happens.

Evaluation may be based on student achievements and on the views of various stakeholders, such as students, staff, external examiners and employers. Key questions are: is the strategy working? How has it made a difference to the quality of assessment in the department? Does it need improving? What support is needed to improve it?

As a result of evaluation, a review of the management of the strategy may be needed, or one might need to adjust the strategy. Also loops need to be closed: if external examiners or students have commented, there may have been action, but they need to be told. Note that this re-emphasises the importance not of quality bureaucracy but of quality enhancement.

Improving an assessment strategy will help you to improve the quality of learning and teaching too. To emphasise the importance of this aspect, we follow Mutch and Brown in quoting Derek Rowntree:

"If we wish to discover the truth about an education system, we must look into its assessment procedures. What student qualities and achievements are actively valued and rewarded by the system? How are its intentions and purposes realised? To what extent are the ideals, aims and objectives professed by the system ever truly perceived, valued and striven for by those who make their way within it? The answers to such questions are to be found in what the system requires the students to do in order to survive and prosper. The spirit and style of student assessment defines the de facto curriculum." (Rowntree, 1987)

These remarks apply as much to MSOR as to any other discipline area. There is one final point. The student world is now one in which, while some students continue to learn for the love of it, many more are driven by hurdles and the desire to achieve. When most students must have (sometimes quite substantial) part-time jobs and are thus in some sense part-time at university, they will inevitably become strategic in the way they manage their time, and it will become even more so the case that what you assess is what you get. Evaluation and enhancement of the assessment process is thus central to the quality of what we do, and the process of self-examination involved in creating and maintaining a strategy can only help.

Appendix: A checklist of questions and comments

This checklist is reproduced with some small modification from Mutch and Brown (2001), as an aide-memoire for those given the task of drafting a strategy. All questions contain implicit values, and there are no ideal solutions. Take into account: degrees of freedom allowed by institution; existing approaches in the department; MSOR traditions; range of students participating in the programme; and how much the department wants to be innovative.

Purposes of the strategy

Generic purpose of the strategy	What is the generic purpose of the strategy? This might be a simple statement along the lines of sustainability, maintenance, improvement etc.
Generic questions	Five useful questions of assessment are: - What kinds of things do you want students to learn? - What opportunities for learning will be used? - What methods of assessment? - What criteria? - How will you evaluate the success of the methods?
Purposes of assessment	What are the purposes of your assessment procedures? Figure 1 provides a starting point.

Programmes and students

Programmes	Different modes and levels of study.
Subject	What are the customary approaches to assessment in MSOR? What are the debates? Do you wish to modify or extend the usual approaches?
External bodies	How do the views of employers and professional bodies affect your assessment?
Students	What is the profile? Are there diverse backgrounds? Do you have mature students? Do assessment procedures take this into account? Are there special provisions for students with disabilities? How to minimise plagiarism?
Levels and student development	Do you want different methods of assessment at different levels (same method but more demanding, or different method)? If the latter (e.g. project or dissertation) how do you prepare students for change?
Assessment loads	What is the load on students *and* staff? Are they comparable across modules? How to get comparability? (Difficult - some use 3 hour paper = 3000 word assignment but does that help in MSOR?)
Use of resources	Must be feasible and practical. How much staff time on preparation, marking, checking, moderation, preparing reports, attendance at meetings? Are there alternative methods which could assess the learning outcomes for less effort? Are physical resources sufficient (e.g. equipment for presentation skills)?

Guiding principles

Intended learning outcomes	Trying to test separately every module learning outcome will lead to over-assessment. Assessment of *programme* outcomes is key, and they should be taught, practised and assessed at every level. Use assessment tasks to assess clusters of learning outcomes. One can use matrices to check coverage.
Coursework versus examinations?	Be explicit about the balance you favour and why.
Diversity or consistency?	Approaches here vary widely. Some begin with more coursework (and more support) at entry and move towards a project/dissertation and more exams at final level. What is your underlying rationale? What is more important: diversity or consistency? Should the approach or rationale be changed? Whatever your answer, it is important to fit methods to learning outcomes to be assessed.
Assessment methods	See Guide for Lecturers. Which methods do you use and why? Are they appropriate to the learning outcomes?
Equity	Consistent marking within and across modules is important. Issues are use of criteria, internal consistency of a marker; consistency within a marking team or across a programme. What procedures do you need? Are they working? What analysis of student performance across modules?

Staff/educational development	What support or training is provided for new staff, part-time staff and teaching assistants?
Minimising discrimination while maintaining standards	Some modes of assessment have a gender bias (E.g. Men often do better in multiple choice tests and women in coursework) Students with disabilities have special needs. Check whether your methods have differential effects on different groups - light blue touch paper to start the debate, and enjoy the controversy ....
Helping students learn	Feedback is central to helping students to learn. What kind of feedback do you provide (e.g. rating schedules, marking schemes, written comments, tutorials, and overviews of the performance of the class on the assignment? Are overviews provided in class or on the web? Does the feedback improve student learning? Are students taught how to use feedback? Is the use made of this feedback monitored? If reflection is a valued characteristic, how is it taught and how do you provide feedback on levels of reflection?

Procedures

Design of assessment tasks and examination questions	What procedure do you have to ensure the tasks and examinations fit the learning outcomes? When are the examination questions set and submitted for approval?
Timing of information given to students	When are students given the assignments, the date of submission and the examination timetable? Are students asked to keep a copy of work submitted? Are these tasks published, for example, in the student handbook?
Marking times and feedback?	What is the timetable for marking assignments and examination scripts? Is there sufficient time for marking assignments and examination scripts and for moderation? To whom do students hand their assignments, and is this recorded? Are students always given written feedback, and does the feedback help them to learn? How do you ensure high quality feedback from all staff? Are the systems working?
Criteria	Do all assignments have criteria? What kind of criteria are used in the department? Do the students know and understand the criteria and marking schemes? Are the criteria published? What is the rationale underlying use of your criteria? How do the criteria relate to levels of progression and learning outcomes?
Moderation	How are assignments and scripts moderated? What methods do you use? Despite the fashion of “double blind mark everything”, there is no evidence that the process is any more reliable than sampling based on double marking. Double blind marking of assignments is often not practical or useful for developmental purposes. Is there sufficient time for moderation?
External Examiners, Examining Boards and record keeping	What are the roles of External Examiners and Boards of Examiners? Could the system be improved? Is there an induction process for new examiners? Who organises papers for meetings of boards? How? Where and how are records kept? Are External Examiners informed of any actions taken (or not taken) as a result of their report? Do tutors have easy access to records of student achievement? Are there appropriate data systems in place? Are they consonant with the university system?
Appeals	What is the appeals procedure? Is it time consuming? Could it be improved?
Plagiarism	What is the current procedure for handling plagiarism? How could plagiarism be minimised in the department?
Reporting procedures	How is assessment strategy monitored and evaluated? What are the roles and responsibilities of members of departments and committees? What is the flow line of reporting procedures? Does it work?
Staff development	What support will be provided to staff for the development and implementation of the strategy? Administrative and technical staff, as well as academic staff and teaching assistants may need mentoring, briefings and workshops. Who will provide them and when?

References

Beevers C and Patterson J (2002) "Assessment in Mathematics" in Kahn P and Kyle J (editors), *Effective Learning and Teaching in Mathematics and its Applications*, London: Kogan Page, pp 49-61

Haines C and Dunthorne S (1996) (editors) *Mathematics Learning and Assessment: Sharing Innovative Practices* London: Edward Arnold

Holton D (2001) (editor) *The Teaching and Learning of Mathematics at Undergraduate Level: an ICMI Study*, Dordrecht: Kluwer Academic Publishers

Houston K (2001) "Assessing undergraduate mathematics students" in Holton D (editor), *The Teaching and Learning of Mathematics at Undergraduate Level: an ICMI Study*, Dordrecht: Kluwer Academic Publishers, pp 127-136

Kahn P and Kyle J (2002) (editors) *Effective Learning and Teaching in Mathematics and its Applications*, London: Kogan Page

Mutch A and Brown G (2001) *Assessment: A Guide for Heads of Department,* LTSN Generic Centre Assessment Series No 2, York: LTSN

QAA (Quality Assurance Agency) (2000) *Code of Practice: Quality and Standards in HE*, http://www.qaa.ac.uk/public/COP/COPaosfinal/contents.htm (accessed 31 January, 2004)

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

Assessment in Mathematics, Statistics and Operational Research

Guide for Lecturers

Introduction

This guide draws heavily on the corresponding booklet in the Generic Series, Booklet 3, written by George Brown (Brown, 2001).

Probably the most useful piece of advice a *Guide for Lecturers* could give is to say, "First of all write the *learning outcomes* of a module *very carefully*." Thoughtful learning outcomes are the key to good assessment. But the next piece of advice would be, "Read the entire guide before starting." This is the thrust of Assessment Series Booklet 3, written by George Brown, recently retired from the University of Nottingham. He stresses the point that alignment between learning outcomes, methods of learning, assessment methods and assessment tasks is crucial.

But the next piece of advice would be, "Read the entire guide before starting"

Hence the lecturer, as teacher and assessor must know about all of these topics in order to perform well. Good MSOR lecturers with many years of experience will probably have had implicit learning outcomes in mind, they will know how students learn and how to facilitate that learning. They will know how to set three hour, unseen, written, examination papers that are reliable and discriminating. But given the current demands from our students, for transparency in everything to do with assessment and relevance in everything to do with the curriculum, they, along with everyone else, must meet these demands. This is a good thing, because it requires the introduction of different methods of assessment, methods that are better than written examinations at assessing some learning outcomes like, "On successful completion of this module students should be able to demonstrate effective teamwork skills."

The MSOR community in the UK is well aware that Subject Benchmark Statements have been written and published by the Quality Assurance Agency (QAA, 2002). These statements each describe their subject and list, as learning outcomes, the "benchmark" standards expected of graduating students at threshold honours level and at modal honours level. While the benchmark statement mainly relates to final year achievement, clearly the foundation for this must be laid in first year and built upon in succeeding years. This section includes a discussion of benchmarking.

The section starts with a reference to the Briefing number 1 on key concepts and moves on to a discussion of alignment. It deals with the writing of learning outcomes, including the influence of the benchmark statement. It discusses methods of learning. It looks at the purposes and principles of assessment and catalogues many methods of assessment, commenting particularly on those that seem to be most appropriate for MSOR.

Bear in mind that with this generation of students, "what you assess is what you get". Assessment is perhaps the most important function that lecturers perform and therefore it is important that it is carried out professionally.

Key concepts

The reader is encouraged to read now Briefing number 1: Key Concepts. This explains the ideas associated with formative and summative assessment, norm and criterion referencing, reliability, validity, affordability and usability.

Bear in mind that with this generation of students, "what you assess is what you get"

It should be the aim of every examiner to set examinations and assessment tasks, and to devise assessment schemes that are as reliable and as valid as possible given the constraints of affordability, and to be of use to every stakeholder in the whole process of educating students.

Alignment

Every horse racing aficionado knows that "there are horses for courses". So too there are appropriate learning methods, appropriate assessment methods and appropriate assessment tasks for different learning outcomes, and it is important to align these in a sensible way.

The process of writing a programme specification or a module specification is a bit like doing mathematics - the final presentation does not always reflect the process of getting there. The final presentation of a mathematical idea is usually of the form: - definition - theorem - proof. But the process of getting there is messy and circuitous. The final presentation of a programme or module specification is usually of the form: - aims - intended learning outcomes - content - methods of learning - assessment methods - assessment tasks. Again the process of getting there is messy and circuitous. The design committee will have an idea of what they want, bearing in mind the departmental strategy. (See the Guide for Heads of Department). They will draft overall aims for the programme and write draft learning outcomes. (Learning outcomes are a list of things that will be expected of students when they finish the programme or module. See below.) They will discuss the likely learning methods - lectures, tutorials, practicals, private study, and so on. They will agree on content and the most appropriate assessment methods and assessment tasks for each of the learning outcomes. After debate, the committee will produce a course document specifying the programme and outlining the modules. This will be subject to further revision after experience operating the course, with the goal of obtaining good alignment.

The development of "assessment tasks" will require the writing and revision of assessment criteria and grade descriptors. The purpose of each task, how the whole assessment process will be managed and how feedback will be given to students and to lecturers will be decided. When appropriate, the procedures for giving "feedout" will be put in place. ("Feedout" is a term used to describe the reports of assessment outcomes given to university managers, postgraduate schools and prospective employers.)

Learning Outcomes

Learning outcomes are simply statements of the things it is intended that students will be able to demonstrate or do when they complete the programme or a constituent module. Usually they are written in the future tense and contain an active verb. At programme level they should be subject specific but generic enough so that not more than about six are required. At module level, again there should only be about six; they should be more specific than at programme level and they should relate to one or more of the programme learning outcomes.

Learning outcomes are simply statements of the things it is intended that students will be able to demonstrate or do when they complete the programme or a constituent module

The programme should be designed in such a way that each of the programme learning outcomes is achieved in at least one module, but each module does not necessarily have to achieve all of them.

Learning outcomes may be categorised under:

A. Subject knowledge and understanding;
B Intellectual qualities;
C Professional or practical skills;
D Transferable or key skills;

Here are examples of programme learning outcomes.

On completion of this programme of study, students will be able to:

(i) Demonstrate knowledge and understanding of the foundations, principles, techniques and applications of selected mathematical concepts;
(ii) Construct logical arguments;
(iii) Analyse and solve both routine and complex problems, appreciating the limitations of the techniques employed;
(iv) Construct and use theoretical models of systems and situations, employing a creative and critical approach;
(v) Use information and communications technology to help solve problems and report findings;
(vi) Demonstrate competence in the key skills of communication, teamwork, independent learning and time management.

Learning outcome (i) belongs to category A; learning outcomes (ii), (iii) and (iv) belong to category B; learning outcomes (v) and (vi) belong to categories C and D respectively.

Note that when constructing questions to test learning outcome (ii), it is important to avoid "memory" questions.

The learning outcomes of a programme should always be compared with the modal learning outcomes for Honours Degrees set out on page 14 of the Benchmark Statement (QAA, 2002):-

A graduate who has reached the modal level should be able to:

(i) Demonstrate a reasonable understanding of the main body of knowledge for the programme of study;
(ii) Demonstrate a good level of skill in calculation and manipulation of the material within this body of knowledge;
(iii) Apply a range of concepts and principles in loosely-defined contexts, showing effective judgement in the selection and application of tools and techniques;
(iv) Develop and evaluate logical arguments;
(v) Demonstrate skill in abstracting the essentials of problems, formulating them mathematically and obtaining solutions by appropriate methods;
(vi) Present arguments and conclusions effectively and accurately;
(vii) Demonstrate appropriate transferable skills and the ability to work with relatively little guidance or support.

There are other classifications of learning outcomes, with perhaps the most widely used being that devised by Bloom (1956). He classified educational objectives in the cognitive domain in six hierarchical classes:

1. Knowledge - the recall of information;
2. Comprehension - the translation, interpretation or extrapolation of knowledge;
3. Application - the application of knowledge to a new situation;
4. Analysis - to break down knowledge into parts and show relationships among the parts;
5. Synthesis - bringing together parts (elements, components) of knowledge to form a whole and build relationships for new situations;
6. Evaluation - making judgements about the value of material and methods for given purposes.

Moving from 1 to 6 makes increasingly greater cognitive demands on students.

For each of these six categories, there are appropriate verbs to use when writing learning outcomes and some of these often used in MSOR contexts are given here: -

1. Knowledge - define, describe, name, recall, state;
2. Comprehension - classify, discuss, explain, translate, interpret, paraphrase;
3. Application - apply, solve, use;

4. Analysis - analyse, categorise, compare, contrast, distinguish;
5. Synthesis - create, construct, formulate, design, organise, set up;
6. Evaluation - appraise, evaluate, assess, predict.

Modules offered in the final year of an honours degree programme should be using verbs from categories 5 and 6 more often than those from categories 1 and 2 in their statements of learning outcomes.

Methods of learning

The university or college will provide learning opportunities and learning resources for students. It will provide lecturers who will inspire and inform through lectures, tutorials, practicals, written materials and one-to-one meetings, and who will assess both formatively and summatively through a range of assessment methods and tasks. It will provide libraries and space for students to meet and to study. It will provide computers to use as calculating aids, as tools for communication and as access vehicles to Internet and Intranet resources. One can, of course, only take a horse to the water; getting it to drink is another matter; the horse has to want to drink; it has to be thirsty.

Thankfully most students are thirsty for knowledge. Many are curious and seek understanding. But some want to get a degree by doing as little work as possible. Students take different approaches to learning at different times. Sometimes they will take a surface or superficial approach, seeking only to remember enough facts to answer enough of the examination questions to pass. They will soon forget what they learn in this way, but that may be OK. Sometimes they will take a deep approach to learning, seeking to make sense of their new knowledge, to relate it to what they already know, to modify their concepts to accommodate the new material, to apply their new understanding to new problems (Ramsden and Entwhistle, 1985; Marton and Saljo, 1976).

Students take different approaches to learning at different times.

The implication for lecturers in higher education is that they should aim to promote conditions in which a deep approach to learning will flourish (Gibbs, 1992). Active learning is a means of promoting a deep approach. This encourages students to become more responsible for their own learning through self-initiation and self-appropriation of knowledge. It requires a constructivist approach whereby the student reconstructs, transforms and transfers knowledge to achieve personal understanding. In other words, they make it their own. The elements of active learning are doing - active engagement rather than passive receptivity - reflecting, synthesising, abstracting and evaluating. And, as Briefing number 6 on Self, Peer and Group Assessment indicates, peer tutoring is a form of active learning, which can help achieve deep learning.

It requires a constructivist approach whereby the student reconstructs, transforms and transfers knowledge to achieve personal understanding.

The Warwick Analysis Project (Alcock and Simpson, 2001) is a good case study of an active learning experiment. First year students study analysis - traditionally a "hard" subject - by working through a carefully structured sequence of questions in a text book by Burn (1992). Quoting Alcock and Simpson, "The text consists mainly of questions which develop rationales for the main definitions, construct the central arguments that lie behind the main theorems and allow the students to use those theorems in subsequent arguments." It was found that the students on the pilot scheme scored higher marks on average than students in the control group who had been taught in the traditional way. These authors also suggest that the new course "allows a much more rapid adoption of a new way of thinking required for university mathematics."

The learning outcomes of an MSOR programme are required by the Benchmark Statement to include the learning of transferable skills. A discussion of suitable learning methods for this purpose is given in Briefing number 3: Key Skills.

Purposes of assessment

Brown (2001) states that the three main purposes of assessment are:

1. To give a licence to proceed to the next stage or to graduation
2. To classify the performance of students in rank order
3. To improve student learning

To these maybe added:

4. To inform teachers of the strengths and weaknesses of the learners and of themselves so that appropriate teaching interventions may be employed
5. To inform other stakeholders - society, funders, graduate schools, potential employers
6. To encourage learners to take a critical-reflective approach to everything that they do, that is, to self-assess before submitting

The first two of these together with number 5 could be described as assessment for the purpose of reporting student achievement to themselves, to university managers and to the world. Summative assessment is required for these purposes. Formative assessment is required to improve student learning. Brown writes, "These purposes may overlap or conflict. A common error is to use an assessment task for one set of purposes and then to assume that the results from it are appropriate for other purposes." The MSOR community regularly uses some assessment tasks for both summative and formative purposes. Weekly homeworks and intra - semester class tests are set, marked and returned to students. The marking should be sufficiently detailed so that students can see their strengths and weaknesses and learn from the experience. Marks are

recorded, to chart student endeavour and progress, and to contribute to the aggregate coursework mark for the module. There are arguments for and against using these two tasks for summative purposes. Clearly they are an aid to learning, albeit a costly one in terms of the lecturer's time. The questions are usually fairly straightforward and test core knowledge. If some students are learning quickly the marks may be high and bunched together, not discriminating between students. These students will not learn much from their few mistakes but will have the satisfaction of knowing that their learning is progressing. If some students are learning more slowly, their marks may be low; they may not have grasped pre-requisite knowledge. As a result these students will have learning opportunities and should soon catch up. Either way, if these marks were used for summative purposes, they would not necessarily reflect a student's abilities at the end of the semester. On the other hand, if the marks are not summative, then some students may not bother to do the homeworks unless some other incentive is provided.

When "licence to proceed" is a key concern, then assessment tasks should concentrate on essential, core knowledge and skills, and the pass threshold should be set at a high level. The results of this assessment should not be aggregated with other results, but should be reported separately, and there should be a requirement to pass. Other assessment tasks, like written examinations, should be used to discriminate between students in order to place them in rank order.

Principles of assessment

Brown (2001) lists six principles which he considers to be key:

1. Assessment shapes learning so if you want to change learning then change the assessment method
2. Match the assessment tasks to the learning outcomes
3. Match the criteria to the task and the learning outcomes
4. Keep the criteria simple
5. Be fair, reliable and valid in your marking
6. Provide meaningful, timely feedback

These are fairly self-explanatory.

Methods of assessment

Brown (2001) lists 23 different methods of assessment. Many of these are not appropriate for MSOR so this section contains a catalogue of methods of assessment used in an MSOR community somewhere in the world. It is probably not exhaustive!

Timed, unseen written examinations (closed book).

These are the staple diet of the MSOR community. They are very reliable in assessing particularly the lower order cognitive skills, but not useful for assessing extended investigations. Quite often considerable thought and time are required to construct a fair examination paper which gives all students an opportunity to show at least some of what they know and can do and which discriminates between students of different abilities. The questions may be structured into several sub-questions, which are often related and sequential, with the latter parts having the so-called "sting in the tail" to challenge the very best students. It is now customary to inform students how many marks are available for each sub-question.

They are very reliable in assessing particularly the lower order cognitive skills, but not useful for assessing extended investigations.

It is important to write specimen solutions, and a detailed marking scheme. This helps ensure that the question can in fact be answered in an appropriate time, it helps the internal moderator and the external examiner to vet the paper, and it facilitates the marking, whether by a single marker or a team of markers. See also the section on Norm Referencing in Briefing number 1: Key Concepts.

In the interests of openness and transparency, it is now customary to give students access to previous examination papers, and some institutions also give access to the worked solutions of these previous papers. Giving solutions might be inclined to inhibit some students from trying hard to work the paper; there is a temptation to look up the answer at the first sign of difficulty and consequently students will not learn as much as they could if they persevered. Furthermore, if the structure of the paper is to be changed in any way, be warned that today's students will not take kindly to this if they are not fully informed beforehand!

It is important to have the paper moderated or vetted by a colleague. This helps ensure that the paper is of a good standard and it helps to eliminate errors. And note that external examiners like to receive papers that are error free!

Timed, unseen written examinations (open book).

These are becoming more popular and come in several varieties. Sometimes a student is permitted to bring in to the examination and use any notes and textbooks he or she likes. Sometimes the student may bring in and use only a small number of pages of hand-written notes. The main difference between open book and closed book examinations is that questions requiring simply recall of knowledge should not be set, thus allowing the setting of more questions at higher cognitive levels. Students like the "comfort blanket" in case they forget something, but in practice, it is better for students to have the knowledge in their heads as it is quicker to access it there than in a large collection of books and notes.

Other variations on written examinations

Some people have used "seen" examinations wherein students are given the questions some time in advance. They prepare their answers and then write the answers in the examination hall without the aid of notes. The case for this mode of examining is that it removes the pressure of the time constraint; students can tackle the questions over a lengthy period, come to a good understanding of the

solutions and then only have to remember them for the examination. The main problem is that someone else might write solutions for a student.

The case for this mode of examining is that it removes the pressure of the time constraint; students can tackle the questions over a lengthy period.

Some people have conducted traditional examinations in a computer laboratory so that students could have easy access to all the mathematical software they had been using through the semester. This means that questions could be set which required use of the software and that the examination would more faithfully reflect the learning outcomes of the module.

Individual Project Work

Final year projects have been required of honours degree students by many universities for a long time. This can be a substantial piece of independent work, contributing usually about one sixth of the total work of the final year. The topics set for investigation can be quite demanding and give plenty of scope for student initiative. The project is an opportunity for a student to conduct research, to write up their work in a report and, often, also to present it in a seminar or on a poster or in a *viva voce* examination.

The supervision and assessment of individual project work can be very costly in terms of staff time especially when big cohorts of students are involved. Thus, while such projects are highly desirable, they may not be affordable. Assessment criteria should be devised and a useful source of information on this is the set of booklets edited by Haines and Dunthorne (1996). This pack includes lists of assessment criteria for several different types of projects – pure mathematical investigations, mathematical modelling, statistical projects and more general ones involving areas like the history of mathematics and mathematical educational research. It also includes assessment criteria for different modes of communication – written reports, oral presentations and poster presentations. When the assessment criteria have been drafted then staff have to practice using them, testing each other until there is good agreement on standards and conventions. Haines and Houston (2001) present a case study that illustrates how this procedure might be organised.

Group Project Work

Group project work is often introduced at an earlier stage of a programme, and again it provides opportunity for students to be introduced to the "way of life" of a professional mathematician. It encourages investigation, research and communication, and it has the added ingredient of group work. Working with other people is an important key skill and while students may not always like working in this way, it is a useful learning experience for them. They see at first hand some of the human problems of professional life. The assessment of group work is also discussed in Haines and Dunthorne (1996). One way – possibly the best – to get information about the internal working of the group is to use confidential self and peer assessment. This allows students to comment on how well they believe that they themselves and their colleagues have contributed to the work. Another way is for the lecturer to observe the group at work, although there is a danger that the process of observation will distort that which is being observed. It does have the advantage that sometimes a crisis can be averted by a timely intervention.

One way – possibly the best – to get information about the internal working of the group is to use confidential self and peer assessment.

Ultimately, though, the success or otherwise of the group will depend on the outcome of the assessment of their product – the written report, the oral presentation, or whatever. This mark could be given to each student in the group and it may be modified to take account of the assessment of the working of the group, either by student self and peer assessment or by lecturer observation.

Comprehension Tests

Many subjects use comprehension tests. The methodology is to require students to read a paper and then to set them an unseen written quiz which interrogates their understanding of what they have read. The paper may be quite short, such as some of the short articles that appear in the *Mathematical Gazette*, published by the Mathematical Association, and then it may be given to students for the first time in the examination. Or it may be a longer paper, which would then be given to students some time in advance of the examination. It is good if papers already published in the literature can be used, as it is then an authentic example of professional life. Questions can be asked, for example, which require students to expand the abbreviated mathematical working reported in the paper, to evaluate the assumptions made by the author, and to comment on how they might have tackled the same problem. Preparing for a comprehension test when the article is given out in advance encourages students to read critically and reflectively, to try to get into the mind of the author, and to think deeply about the topic. Students should be informed of the purposes and methodology of the test and encouraged, as part of preparation, to discuss the article with colleagues and to try to anticipate the questions that will be asked (Houston, 1993).

Diagnostic Tests

A diagnostic test is a straightforward test of basic knowledge that is usually used when a student starts a programme or module, to ascertain which mathematical topics, if any, the student may have forgotten. The test then indicates a programme of material with which the student may review these topics. Many people are working on the development of computer aided diagnostic tests and remedial teaching software. The LTSN Maths, Stats and OR Network is sponsoring a project to develop a national repository of

diagnostic tests for the use of the MSOR community (see http://ltsn.mathstore.ac.uk).

Computer Aided Assessment

As is mentioned in Briefing number 2: Assessment of Large Groups, computer aided assessment (CAA) is proving useful in this context. It is also proving useful in programmes which include "learning at a distance" or "E-Learning" modules. When first introduced, CAA could only handle relatively simple things like multiple-choice questions. It still has limited use but the technology is advancing all the time, allowing the use of more sophisticated assessment questions. Beevers (2003) edits an on-line journal, which publishes a new article on CAA every month. Publication started in July 2001.

Assessment of Key Skills

See Briefing number 3.

Assessment of Portfolios

See Briefing number 4.

Over the last 10 to 15 years some educational researchers have turned their attention to the learning of mathematics in higher education institutions.

Journal Writing

Student journal writing through a module can be used to help diagnose learning difficulties and to address these at an early stage. Students may be given time at the end of a teaching session to reflect on their learning during that session and to write down their thoughts and feelings, their worries and concerns, what they feel they have learnt and what they are having difficulty with. Or they may be asked to do this overnight, thus allowing some time to digest the day's work. The journals should be read frequently by the lecturer so that formative feedback can be given in good time and appropriate teaching interventions introduced. It can also be quite therapeutic for the lecturer to write his or her own journal!

Further reading

Over the last 10 to 15 years some educational researchers have turned their attention to the learning of mathematics in higher education institutions. Theories about student learning at this level have been proposed and tested. It is helpful to the MSOR lecturer to be aware of this work as it can inform teaching and assessment. The book *Advanced Mathematical Thinking*, edited by Tall (1991) has a good collection of articles which give insights into some of the difficulties students have learning mathematics at this level.

Another useful book is the ICMI Study, *The Teaching and Learning of Mathematics at University Level*, edited by Holton (2001). This has articles on the present day context of mathematics teaching, current and innovative practices, research case studies, using technology, assessment, and the training of school teachers. The article by Houston (2001) contains an extensive bibliography, which could be updated to include interesting articles and books published since then.

A more recent book is *Effective Learning and Teaching in Mathematics and its Applications*, edited by Kahn and Kyle (2002). This has a good selection of articles covering most issues of importance today. There is an article on assessment by Beevers and Paterson (2002) which includes ideas on computer-aided assessment and Hibberd's article includes the assessment of mathematical modelling (Hibberd, 2002).

There are also a number of articles on the assessment of mathematical modelling in the ICTMA books (www.infj.ulst.ac.uk/ictma), the most recent of which are edited collections by Ye et al (2003) and Lamon et al (2003).

References

Alcock L and Simpson A (2001) "The Warwick Analysis Project: Practice and Theory" in Holton D (editor), *The Teaching and Learning of Mathematics at University Level*, Dordrecht: Kluger Academic Publishers, pp 99-111.

Beevers C (2003) http://ltsn.mathstore.gla.ac.uk/index.asp?cat=137 (accessed 30 August 2003).

Beevers C and Paterson J (2002) "Assessment in Mathematics" in Kahn P and Kyle J (editors) *Effective Learning and Teaching in Mathematics and its Applications*, London; Kogan Page, pp 49-61.

Bloom B S (1956) (editor) *Taxonomy of Educational Objective: Book 1, Cognitive Domain,* New York: Longman.

Brown G (2001) *Assessment: A Guide for Lecturers*, LTSN Generic Centre Assessment Series No 3, York: LTSN.

Burn RP (1992) *Numbers and Functions: Steps into Analysis*, Cambridge: Cambridge University Press.

Gibbs G (1992) *Improving the Quality of Student Learning*, Oxford: Oxford Centre for Staff Development.

Haines CR and Dunthorne S (1996) (editors) *Mathematics Learning and Assessment: Sharing Innovative Practices,* London: Edward Arnold.

Haines CR and Houston K (2001) "Assessing Student Project Work", in Holton D (editor) *The Teaching and Learning of Mathematics at University Level*, Dordrecht: Kluger Academic Publishers.

Hibberd S (2002) "Mathematical Modelling Skills" in Kahn P and Kyle J (editors) *Effective Learning and Teaching in Mathematics and its Applications*, London; Kogan Page, pp 158-174.

Holton D (2001) (editor) *The Teaching and Learning of Mathematics at University Level*, Dordrecht: Kluger Academic Publishers.

Houston SK (1993) "Comprehension Tests in Mathematics: I and II", *Teaching Mathematics and its Applications,* **12**, pp 60 – 73 and 113 – 120.

Houston K (2001) "Assessing Undergraduate Mathematics Students" in Holton D (editor) *The Teaching and Learning of Mathematics at University Level*, Dordrecht: Kluger Academic Publishers, pp 407-422.

ICTMA website, http://www.infj.ulst.ac.uk/ictma/books.html (accessed 29 August 2003).

Kahn P and Kyle J (2002) (editors) *Effective Learning and Teaching in Mathematics and its Applications*, London; Kogan Page.

Lamon SJ, Parker WA and Houston SK (2003) (editors) *Mathematical Modelling: A Way of Life: ICTMA 11*, Chichester: Horwood Publishing.

Marton F and Saljo R (1976) "On Qualitative Differences in Learning" *British Journal of Educational Psychology,* **46**, pp 4-11.

QAA (Quality Assurance Agency) (2002), *Benchmarking Academic Standards*, http://www.qaa.ac.uk/crntwork/benchmark/phase2/mathematics.pdf (accessed 19 June 2003).

Ramsden P and Entwistle N (1985) "Student Learning Research: Retrospect and Prospect' *Higher Education Research and Development,* **4**, pp 51-69.

Tall D (1991) (editor) *Advanced Mathematical Thinking*, Dordrecht: Kluger Academic Publishers.

Ye Q, Blum W, Houston SK and Jiang Q (2003) (editors) *Mathematical Modelling in Education and Culture: ICTMA 10*, Chichester: Horwood Publishing.