ltsn generic centre
Learning and Teaching Support Network

Assessment Series No.12

A Briefing on Assessment of Large Groups

Chris Rust

Chris Rust is Head of the Oxford Centre for Staff and Learning Development and Deputy Director of Human Resources at Oxford Brookes University. He was Course Leader for the University's initial training course for new teaching staff for six years, and along with three colleagues he helps to provide both staff and educational development support to the University's academic Schools and support Directorates. Through the Oxford Centre for Staff and Learning Development he contributed to the design and delivery of a national programme of staff development in higher education on the issue of teaching more students. He has also run numerous workshops in other institutions in the UK, and some internationally, on a range of other topics including assessment strategies, developing teaching in higher education, and course design and evaluation. He has researched and published on a variety of issues including: the experiences of new teachers in HE, the positive effects of supplemental instruction, ways of diversifying assessment, and the effectiveness of workshops as a method of staff development. He is a Fellow of the Staff and Educational Development Association and an ILT member and accreditor.

Contents

Generic Centre Guides and Briefings

Welcome to the Learning and Teaching Support Network Generic Centre's series of Assessment Guides and Briefings. They aim to provide a series of overviews of important issues and practices in the field of assessment for the higher education community.

The Assessment Guides are intended for colleagues with particular roles and for students, as their titles suggest. The Briefings are primarily intended for lecturers and other staff involved in supporting learning.

The Assessment Series is a snapshot of a field in which development is likely to be rapid, and will be supplemented by specific case studies produced by the LTSN Subject Centres.

The series was developed by Brenda Smith and Richard Blackwell of the LTSN Generic Centre with the support of Professor Mantz Yorke. Experts in the field were commissioned for each title to ensure that the series would be authoritative. Authors were invited to approach the issue in their own way and no attempt was made to impose a uniform template.

The series editors are grateful to colleagues in LTSN Subject Centres and other senior colleagues who refereed the series, and of course to the authors for enabling its publication.

We hope that you will enjoy the Assessment Series and find it interesting and thought-provoking. We welcome your feedback and any suggestions you may have for future work in the area of assessment.

Professor Brenda Smith
Head, LTSN Generic Centre

Richard Blackwell, Senior Adviser,
LTSN Generic Centre

Professor Mantz Yorke,
Liverpool John Moores University

November 2001

<u>Summary</u>

This paper identifies the major assessment issues of larger classes - that it is likely to be done less well and/or less often - and the resulting negative effects on student learning and achievement. It argues that assessment strategies should be a major part of learning and teaching strategies at both institutional and departmental level, and offers a number of specific strategic responses which might be used to overcome the above issues.

It further argues that disciplines need to be prepared to look for solutions beyond their traditional cultures and practices, and then offers detailed examples of possible solutions grouped under the following six headings:

1. Front-ending

2. Do it in class

3. Self- and peer-assessment

4. Group assessment

5. Mechanise the assessment

6. Strategic reduction

Finally the paper describes five individual case studies from a range of disciplines which exemplify the implementation of some of the above strategies.

Issues

According to Universities UK (formerly the Committee of Vice-Chancellors and Principals) in their report "New Directions for Higher Education Funding, reported in the Guardian newspaper, "Since 1989 resources per student have fallen by 38%, following a decrease of 20% between 1976 and 1989. Staff-student ratios declined to an average of 1:17 (1:23 if funding for research which is included in the average unit of funding is excluded) (Macleod, 2001).

As class sizes go up, assuming the staffing remains the same and if the same assessment methods continue to be used, one of two things is likely to happen with regard to assessment, and in both cases the effect on students is detrimental. It is either likely to be done less well - less rigorously and with less and more superficial feedback to the student - and to take longer to be returned. Or, the amount of assessment on the course will be reduced - and it is likely to be the formative assessment that is reduced (or even done away with completely) first. In both cases, tutors are less likely to be available and subsequently there is less opportunity for contact with the tutor to discuss their comments after the work has been assessed.

In both cases, a major casualty is the amount and quality of feedback given to the student, with predictable negative effects:

• With less detailed feedback, it is much harder for students to see what their strengths and weaknesses are and how they can improve in the future. We can therefore expect a decline in future performance, and if this is repeated this can lead to a continuing spiral of poorer performance.

• With longer gaps between the submission of work and the receipt of feedback, students are less likely to pay attention to or be interested in the feedback, even if it is of a reasonable quality, with effects on future performance similar to that described above.

• With less opportunity to discuss and clarify tutor comments, students are unlikely to improve their performance.

• And with fewer formative assessment opportunities, the summative piece of work receiving the feedback is likely to have been done less well in the first place.

Possibly even more important than the issue of feedback, however, is the potentially detrimental effect of less assessment. It has long been recognised that probably the biggest influence on a student's approach to their studies is the assessment regime of the course (Rowntree, 1977; Ramsden, 1992; Gibbs, 1992; Brown, 1997; et al):

• If less of a course is sampled through assessment, students may very easily become selectively negligent, only studying in depth those parts they think/know will be assessed.

• If with increased class size there is a reduction in personal tutor/student contact, assessment is even more likely to become the prime motivator for many students. If there is also less assessment, motivation may drop and only pick up sporadically when the next assignment is set.

Everything said so far has been on the assumption that staffing has not increased to match the increased number of students but it should be noted that even if staffing has increased this can also bring its own set of problems. If the number of assessors for a given piece of work increases this will inevitably increase the problems of maintaining marker reliability and consistency.

The creative solutions and case studies that follow (see 5 and 6 below) all suggest possible ways of addressing one or more of these problems.

Learning and Teaching Strategies

Given the acknowledged importance of assessment in shaping the students' approaches to learning (as already mentioned above), and the increased importance put on assessment by the QAA's new quality framework, assessment should be a major consideration in any learning and teaching strategy.

With regard to assessing larger groups, the important thing to recognise is that there may well be strategic solutions which can only be implemented at departmental or even institutional level, and which are beyond the control of individual tutors. One example of this is the university that has made it a strategic policy to have as much first year assessment as possible transferred to computer-marked objective tests in order to release staff time for a greater focus on research activity. They have consequently put in the necessary expenditure on hardware, software, and supporting infrastructure to make this possible.

A key area at institutional level which may well benefit from review is the university's examination and assessment regulations, and strategic amendments to these regulations might well be possible which can reduce the assessment burden on tutors. For example, you can:

• Revise regulations devised for an earlier time (e.g. there should be two three-hour unseen exams; all work should be double-marked; essays should be 5,000 words) to ensure they do not lead to inappropriate and unworkable assessment arrangements.

• Streamline arrangements for re-sits which often do not need to be as elaborate as the first assessment because they only need to demonstrate threshold competence and are usually not graded

• Allow for the use of 'course requirements'. If you want students to undertake work but do not have the time to mark it yourself, there is every chance they will not do it. Make sure the regulations allow you to make such work a 'course requirement' and a pre-requisite for sitting the exam.

It may be possible to introduce some strategies at a department level without changes to the university's regulations:

• Set clear deadlines for work to be submitted and do not accept late submissions which would extend the period you have set aside for marking

• Set strict word limits and refuse to read material beyond these limits. It is good practice anyway for students to learn to work to word limits and to time.

• Require all work to be handed in to administrative staff or technicians and insist that it is checked off when submitted. Keep as far away from the administration of handling assessed work as possible.

• Require all work to be word-processed (providing students have sufficient access to computers and to training, and students with genuine disabilities are not unfairly penalised).

The easy and obvious solution to the problem of assessing larger classes is to reduce the amount of assessment but if this is done in an ad hoc and piecemeal way it may have very damaging consequences on student learning as has already been explained. But if the assessment for the complete student programme is reviewed it may well be possible to identify aspects of the course which are over-assessed, and assessments which can be changed, reduced or even done away with, with no harmful effect at all.

A more limited example of a specific departmental strategy is the Biology department who were concerned that on modules with large student numbers it was not possible for the module tutor to find sufficient time for individual student tutorials to discuss the feedback given after the assessed work had been returned. Their strategic solution to this was to introduce a department-wide policy that all first year assessed work (the year where most large classes are found) is to be collected from the student's personal tutor rather than the module tutor. This has not only had the effect of sharing the load across the department but has also helped to strengthen the personal tutor system and personal tutor/student relationship.

Disciplinary differences

When it comes to assessing larger groups it is possible that the problems and issues may vary in different disciplines, and therefore so too may the appropriateness of possible solutions. For a start, what is considered a large group may vary considerably from one discipline to another. In art and design, for example, where traditionally great emphasis has been put on ongoing feedback through the one-to-one tutorial in the design studio even a small increase in numbers can have a dramatic effect.

The most important difference, however, is probably that different disciplines have different assessment cultures and traditions. Some discursive disciplines may be almost totally reliant on the use of academic essays, some under exam conditions, for example, while in others essays may not be used at all. If creative solutions are to be found to the problems described above lecturers need to ask some fundamental questions about their practices. Do they genuinely need to be as they are because of the fundamental nature of the discipline and what is being assessed, or whether the differences are simply historical and/or cultural (see Gibbs, 2000). Solutions may well be found by looking beyond disciplinary traditions and lecturers need to be open to the possibilities of borrowing and amending practices from other disciplines. While in some disciplines it may be easier to create objective tests and/or model answers, for example, no such possible solution should be ruled out blindly without serious consideration simply because that is not the way it has been done in this discipline in the past.

Creative solutions and examples

The possible solutions detailed below have been grouped into the following six strategies:

1. Front-ending

2. Do it in class

3. Self- and peer-assessment

4. Group assessment

5. Mechanise the assessment

6. Strategic reduction

It should be noted however that these strategies are not mutually exclusive but can overlap and be combined.

Front ending

The basic idea of this strategy is that by putting in an increased effort at the beginning in setting up the students for the work they are going to do, the work submitted can be improved. Therefore the time needed to mark it reduced (as well as time being saved in less requests for tutorial guidance).

Ways of doing this essentially fall into two categories:

i. Full briefing instructions/checklists

This involves anticipating everything the student may do wrong or misunderstand about the content and purpose of the task. Depending on the subject and the task, this may take the form of detailed briefing instructions (see example below) while in other contexts something more akin to a checklist might be more appropriate.

Writing up the method section

If you are writing up an experiment or study begin the method section by describing the general features of the experimental approach, stating the independent and dependant variables and the units of measurement. The independent variable is that which is manipulated by the experimenter, whereas the dependant variable is that which is measured. If you are reporting a study where there are neither independent nor dependant variables, a brief description of the design of the study should be included. This section is then broken down under several sub-headings.

(a) Subjects. State the total number of the subjects used in the study. Then give the number of females and males. Where possible give the age range, average age and the source of the sample. For example, 'The sample consisted of 80 University students enrolled on The Individual and Society module. There were 35 males and 45 females.' In reporting psychological findings it is important to specify your subjects as precisely as possible since the individual differences in performance are likely to produce important differences in results.

(b) Materials. Description and/or diagrams of materials used and apparatus arrangements. In most cases it is necessary only to name the piece of apparatus used since descriptions are available elsewhere, e.g. stopwatch, Eysenck Personality Inventory. Copies of questionnaires may be included in the appendix to your report, although this is not strictly necessary. If, however, the equipment or the questionnaire has been designed specially for the study, a full description is required.

(c) Procedure. Detailed descriptions of what happened during the experiment or study.

As an example of detailed briefing instructions, here is just part of five pages of instructions on how to write up a practical report for a psychology laboratory course with about 100 students:

Report verbatim the instructions which were given to the subjects. The guiding description here is that your description of procedure should be sufficiently detailed to allow another scientist to repeat your experiment in exactly the same way as you performed it. Discuss the provisions made to isolate and control all the relevant variables which may have influenced the results, e.g. practice, fatigue.

This section should conclude with an outline of how the scores were collected and how the results were scored or coded.

ii. Clarification of the assessment criteria

There is disappointing evidence (Price & Rust, 1999; O'Donovan , Price & Rust, 2001) that while giving the students the assessment criteria which are to be used when the task is set is certainly better than not doing so, it may not actually make much significant difference to the subsequent quality of the work. For this to happen, what is needed is for the students to actively engage with the criteria in some way. At a bare minimum, this might involve the tutor going through the criteria, discussing them with the students and explaining what they mean. Far better though (Nelson, 1994; Price, O'Donovan , & Rust, in press) is to get the students using them in some kind of marking exercise, either using samples of work from a previous year's cohort (suitably anonymised and with the authors' permission) or with specially written pieces written by the tutor for the purpose. Through trying to use the criteria themselves and engaging with the work of another the students gain a much greater insight in to what the key words in the criteria actually mean, and therefore what is expected of them. This need only take up the time of one seminar session and in terms of the outcomes will be time very well spent.

Doing it in class

This strategy requires the tutor to look for aspects of assessment that could be done within the allocated class time rather than outside. The possible types of activity can fall into at least three different categories:

i. Giving general rather than individual feedback

If the problem is finding time to give individual feedback to each student on formative work that they have undertaken, one solution may be to take time in class to give general feedback to the whole class, highlighting things which had been done well or badly, common mistakes and misapprehensions, etc. This strategy can be combined with self and/or peer assessment (see below) in that after the general feedback has been given the students can have time (or encouraged to do so after the class) to consider which aspects of that feedback apply to their piece of work and/or that of one or more of their peers.

ii. Setting assignments which can be marked in class

Instead of written assignments, students can be set tasks which lead to their findings being presented in class, either as oral presentations or as posters. As well as saving the tutor's marking time these methods have the additional advantages that they help the students to develop a wider range of communication skills, allow their work to be shared with the rest of the class, and enable them to receive immediate feedback. Peer assessment can also be used (see below) with the possible positive benefits identified above with regard to marking exercises.

iii. Setting assignments which can be undertaken in class

A good example of this is the 'Instant Lab report' (Gibbs, Habeshaw and Habeshaw 1993). Instead of students writing up long lab reports after the lab which can take a long time to mark and enable the student to hide many of their shortcomings, they write a much briefer report in the lab itself and hand it in as they leave. In some cases it may even be possible to mark it there and then before they leave.

Self- and peer-assessment

Students can perform a variety of assessment tasks in ways, which both save the tutor's time and bring educational benefits, especially the development of their own judgement skills.

i. Self-assessment

This could be as simple as going through a checklist (see below, Fig 1) and making it a requirement that the completed checklist must be submitted with the work in order for the work to be marked. Unless the work is literally being done at the last minute, it is a particularly dim student who ticks title missing, for example, rather than adding a title. In requiring students to review their work in this way it can ensure the prior correction of many basic errors and omissions and thus greatly reduce marking time enabling the tutor to focus on the far more important strengths and weaknesses of the work.

An alternative to a detailed checklist is to simply have a small number of more open-ended titles which the students is required to respond to such as:

• strengths of the piece of work

• weaknesses of this piece of work

• how this essay could be improved

• the grade it deserves is...

• why this lab report deserves better than a ... grade

• what I'd have to do to turn this into a ... grade project

• what I'll pay attention to in my next design is...

• what I'd especially like your comments on

ii. Peer-assessment

Possibly easier than assessing one's own work, especially to start with, is to read and comment on the work of others. Students should therefore be encouraged to get into the habit of getting informal feedback from each other, and if instituted formally this may be a very effective way of ensuring the students get feedback which the tutor does not have time to give. It can also have very definite educational benefits.

An example of this is a course in Geography that originally required the students to write two essays, one towards the beginning of the course and one near the end. The tutor became increasingly despondent that despite all the efforts spent marking and writing comments on the first essay invariably there was little improvement in the second and as student numbers were rapidly rising it was becoming increasingly difficult to find the time to maintain the quality of the feedback given.

Instead of two essays the course now requires only one, which is written in two stages. In the first stage, the students write a first draft of their essay by a given date. In a seminar session, they are then paired up and each pair reads and gives detailed feedback on the other's piece of work. In the light of this feedback, the students then redraft their essays. When the essay is finally submitted it is accompanied by an account of how the feedback has been used, e.g. 'I've included more sources because the first draft was criticised for using only two. I've kept the introduction the same even though it was criticised as unclear because I don't agree…' etc.

As a result of this change the tutor has halved his marking load. It is true that only one topic is now assessed rather than two as before but the tutor believes that educationally this is more than compensated for by the following benefits:

• it develops the students critical faculties

• significantly better work is produced

• it is more like the 'real world' - good writing involves redrafting in the light of criticism

Fig 1. Psychology Practical Comment Sheet

Name:	Date Submitted:
Practical:	Mark:
	Marker:

Checklist of Comments

Title
    ( ) Missing ( ) Correct ( ) Incorrect ( ) Vague ( ) Too short ( ) Too long
    ( ) Incorrect but adequate

Abstract
    ( ) Needs the heading "Abstract" or "Summary" ( ) Section missing ( ) Too short
    ( ) Too long (max. 200 words) ( ) Unclear
    ( ) Wrongly placed, it should be at the beginning
    ( ) Omits hypothesis / aim – design procedure results – conclusion
    ( ) Material which is here belongs elsewhere, e.g.
    ( ) Clear ( ) Succinct

Introduction
    ( ) Section missing ( ) Heading missing
    ( ) Too short (min. 300 words) ( ) About right length ( ) Too long (max. 1,000 words)
    ( ) Follows handout too closely ( ) Rambling and unfocussed
    ( ) Does not incorporate a statement of the hypothesis
    ( ) Rationale for study missing
    ( ) Does not review previous empirical findings ( ) Omits relevant readings
    ( ) Does not consider appropriate theories
    ( ) Some material included here belongs elsewhere, e.g.
    ( ) Inappropriate use of references ( ) Well argued
    ( ) Shows set reading has been done

Method
    ( ) The entire section is missing ( ) Should be sub-divided as below: Subjects
    ( ) Number? ( ) Groups? ( ) Sex? ( ) Age? ( ) Naïve to purposes of study
    Materials/Apparatus
    ( ) Section missing ( ) Not enough detail ( ) Too much detail ( ) Needs diagram
    Design
    ( ) Section missing ( ) Control(s)? ( ) Balancing? ( ) Randomisation?
    Procedure
    ( ) Section missing ( ) Instructions to subject? ( ) Details missing
( ) Too detailed ( ) Whole section clear and detailed

Results
Tables
( ) Missing ( ) Summary table needed
( ) Calculations / Raw data go in Appendix
( ) No numbers / titles on tables (e.g. "Table 1: Mean errors for each age group")
( ) Untidy ( ) Neat
Figures/Graphs
( ) Missing ( ) Axes need labelling ( ) Key to symbols? ( ) Wrong items plotted
( ) Bad scaling on axes ( ) No Numbers / titles on figures (e.g. "Figure 1 : Graph of")
( ) Untidy ( ) Neat
Description of Data
( ) Missing ( ) Too short ( ) Good length ( ) Put some of this in Discussion
Statistics
( ) You have not done all the tests described in class
( ) Link this with a table / result
( ) Arithmetic errors ( ) Tables / figures neat and well presented
( ) Verbal description clear – precise
Discussion
( ) Missing ( ) Little evidence that you have done the required reading
( ) Too short ( ) Too long
( ) Mention problems with procedure / design
( ) Does not pick up points raised in the introduction
( ) Conclusion missing ( ) Your conclusion is not justified from the data presented
( ) Material has been included here which would go better in the introduction
( ) Satisfactory ( ) Well organised ( ) Well organised and well argued
( ) Contains some novel and interesting opinions
Reference
( ) Some references are incomplete
Minimum information is:
(a) first author (b) title of article or book (c) title of journal (if relevant)
(d) volume number
(e) year of publication (f) publisher (books only)
( ) Some references made in report are not detailed here
( ) Some references are inaccurate
( ) Section missing ( ) Satisfactory
General
( ) Poor ( ) Fair ( ) Good ( ) Very Good ( ) Excellent ( ) Too brief overall
( ) Too hurriedly written ( ) Report not set out in formal order ( ) Poor spelling
( ) Poor grammar ( ) Untidy ( ) Difficult to follow your arguments: muddled, disorganised
( ) Too long (you need to demonstrate skill in condensing your argument)
( ) Overall presentation above average
( ) Demonstrates reading beyond set references and extra marks have been awarded for this

<u>iii. Peer marking using model answers</u>

This may not be possible in all disciplines but if it is possible to have model answers for work that is set this strategy can have a number of benefits.

One example is an engineering course where students who did badly in the exam were known to be failing because they needed more practice in working through numerical problems. There was no way, however, that the staff could contemplate more marking and if it wasn't assessed how could they get the students to seriously undertake the practice problems?

The solution they adopted was to set the students problems on a regular basis to solve in their own time, and to allow twenty minutes at the start of certain lectures for these to be marked. This is done by rows of students swapping work; the lecturer then leads them through a model solution. Although 170+ students are involved they have become increasingly efficient at doing this and needed less time - only 7 minutes on one occasion. It was made a course requirement that all students had to have attempted 50 problems over the term (in order to be eligible to sit the end of term exam) but marks did not count. It would therefore have been possible for students to only make a perfunctory attempt at each problem and get them all wrong and still be able to sit the exam. But this did not happen. The students did take the problems seriously, presumably because they did not know which of their class mates would be marking their work and they did not want to be 'shown up'. And the performance in the exam has improved staggeringly.

This is because not only are they getting increased practice in undertaking problems, and the associated feedback, but they are also benefiting from:

• seeing the preferred solution, with the weighting of an examination marking scheme explained

• seeing the variety of approaches taken by their peers

• having to judge the degree to which the work of their peers does or doesn't meet the requirements of the marking scheme

Assess groups

If instead of individual pieces of work students are put into groups to collaboratively produce one product between them it is easy to see that the marking burden for the tutor would be significantly reduced. In addition, further savings may result from less time needed for tutorial support (group tutorials rather than individual) and in some cases fewer placement visits being necessary. There is also the strong educational justification that collaborative skills and the necessary interpersonal skills for group working are high on the list of general 'life' skills which we should be developing.

The major problem of course is that group members may not contribute equally, so how are they to be rewarded fairly? There is probably no easy solution to this but there is a range of possible strategies which may go at least some way to addressing the problem:

i. Group mark

The simplest solution may be to accept that it is slightly unfair but to stick with giving everyone in a group the same group mark on the grounds that probably all forms of assessment disadvantage someone and averaged out over a range of different types of assessment there is a kind of over-arching fairness. The students, however, may not be convinced.

If such a strategy is to be adopted it may help to have clearly understood mechanisms in place for how groups can deal with backsliding members. An example of this is a Business course where if the rest of the group can make a satisfactory case to the tutor that a group member is not pulling their weight a 'yellow card' is issued. This has a number of ramifications:

• if they mend their ways by a certain date the card can be rescinded

• if the card is not rescinded, the individual will receive a 5% penalty in their mark i.e. 5% less than the group mark

• if they continue to significantly under perform a further case can be made to the tutor to issue a red card which sacks them from the group and means they have to produce an individual piece of work

Since the introduction of this scheme, a red card has never been used, no one has ever complained that the system is unfair and most yellow cards (and there have not been that many) have ultimately been rescinded.

ii. Individual contracts

In some cases, if the group task has a number of distinct components, it may be possible for each of the students in the group to have responsibilities - in effect, an individual contract. When marking the product each individual can then be assessed separately on the basis of the degree to which they have met their contracted outcomes.

iii. Divided group mark

This strategy is based on the premise that it is the group members themselves who know best about the relative contributions of individual members so they should be responsible for allocating individual marks. If you have a group of four and the assessed product which they have produced is allocated 50% then you simply give the group 4x50=200 marks to divide between them. For this approach to work you will need to have a number of explicit ground rules:

• the students should be clear about this marking system before starting the task

• the dividing of the marks should be justified in detail against the assessment criteria for the task

In addition, you may wish to consider the following:

• no two students may have the same mark (to prevent the group effectively opting out by deciding to give everyone the same)

• no individual mark may be more than plus or minus 10% from the group mark (the amount is clearly your choice but the underlying idea is that to an extent they should take collective responsibility for the group's product).

iv. Peer-assessment of contributions

This strategy is based on the same premise as the divided group mark - that the group members are in the best position to know about the contributions of individual members – but gives them power over less marks. Instead of giving them all the marks to allocate, there would be a common mark given to each of them for the quality of the group's assessed product but this would be out of say 80%. They would each then assess each of the other group members out of the remaining 20%, against already known criteria, and these marks are then combined and averaged out for each individual, making an individual component mark to be added to the group mark.

v. Viva

In a comparatively short time, it is possible to gain insight into the relative contributions of individual group members through a group viva through questions like, "Whose idea was this?" and "What was your particular contribution to this section?". This can then lead to an individual component mark awarded by you, to be added to a common group mark for the group product. Obviously vivas take time, but marking one substantial assignment and conducting a series of brief vivas can still be quicker than marking individual assignments.

vi. Project exam

In this example the fairness mechanism is deferred; the members of the group all get the same group mark for the assessed project. But they all know from the start that in the module exam there will be one or more questions specifically related to the project. Questions like "Explain the concept of... with specific reference to your project" or "What methods can be used to...? Select one of these methods to explain in detail, using your group project as an example." (Gibbs, G., Habeshaw, S. & Habeshaw, T., 1992) will be quite difficult, if not impossible, for individuals who played little part in the project.

Mechanise the assessment

i. Statement banks

It is conventional for lecturers to write comments in the margins of assignments. Speed of marking and restricted space often means that these comments are clipped and ungrammatical. Need more explanation! Wiggly line, wiggly line, relevance? Such comments can seem curt and overly critical from the student's perspective.

A number of courses in a humanities department have now converted to statement-banks. The comments available in a statement-bank can be more supportive and detailed than lecturer comments. For instance, one such bank lists 34 comments for a lecturer to choose from, including quick comments: 'This is great!!! Do more of this' and detailed comments: 'This introduction/conclusion/ section/phrase feels pasted on and disconnected from the rest of the essay. See if expanding on the ideas in the section before or after works better. Also, ask if this section really relates to the essay or if it is a personal comment about the idea you have just presented. If this is unclear, ask me about it.'

Examples of overview comments include: 'I love your writing style and diction (word choice)' 'I think the tone of this essay makes it less effective' or 'The language of this section/essay is not appropriate for the audience/register/ subject, (too informal or too formal).'

Statement-banks can be introduced at any point along a continuum from low-tech to hi-tech. For instance, at the lower level, students may be given a detailed list of numbered statements and the lecturer then writes the relevant number/s in the margin. This can be a lot quicker and more positive to students than the traditional way of providing comments. At the hi-tech level, statements can be 'loaded' on to a PC, lecturers can punch in numbers and the computer can replace numbers with the appropriate comments and print them out.

ii. Feedback sheets

Feedback sheets (sometimes called assignment attachment sheets) enable tutors to give students feedback quickly simply by ticking boxes to indicate how well the student did against each of the assessment criteria, instead of writing lengthy sentences. Students appreciate such forms (see example Fig 2 below) because they are clear and easy to interpret, and because the feedback is explicitly related to the criteria. They can also be used for prior self-assessment by the student simply by making it a requirement for the student to complete the sheet and to hand it in with the work or the work will not be marked. In this case if you tick the same box as the student against an individual criterion does anything else really need to be said? If the student has ticked a different box then that may highlight a need for you to add some explanatory sentence. This can help you to focus your feedback and the time you spend giving it to where it is needed and will be most useful.

Fig 2. Essay Marking Criteria

Knowledge							
Text	deep, thorough detailed	☐	☐	☐	☐	☐	superficial
Author	wide knowledge used in analysis	☐	☐	☐	☐	☐	no knowledge or not used
Genre	wide knowledge used in analysis	☐	☐	☐	☐	☐	no knowledge or not used
Historical and social context	wide knowledge used in analysis	☐	☐	☐	☐	☐	no knowledge or not used
Essay							
Structure	clear, logical structure	☐	☐	☐	☐	☐	confused list
Quotations	correct, purposeful use properly	☐	☐	☐	☐	☐	incorrect, arbitrary use
Other sources	wide range relevant properly	☐	☐	☐	☐	☐	few, irrelevant im-properly referenced
Grammar spelling	correct	☐	☐	☐	☐	☐	many errors
Personal							
Response to text	vivid, personal	☐	☐	☐	☐	☐	little response
Viewpoint	clearly expressed	☐	☐	☐	☐	☐	no viewpoint
Creativity	imaginative, surprising	☐	☐	☐	☐	☐	predictable
Critical theory							
Understanding	clear grasp	☐	☐	☐	☐	☐	little grasp
Use of methods	wide range appropriate use	☐	☐	☐	☐	☐	little or inappropriate use

<u>iii. Objective tests</u>

Objective tests - multiple-choice questions (MCQs), true/false, insert words or data – offer the possibility of considerable time saving because they can be marked quickly and easily, and can even be marked by machines. Some lecturers may be put off the idea of objective testing because they think they can only test the more superficial learning outcomes and factual knowledge. However, it is possible to devise objective tests that involve analysis, computation, interpretation and understanding. You also need to ask yourself what the purpose is behind the particular assessment task. If your reason for including an end of module exam is to ensure that the students read widely across the subject, then a multiple choice exam would serve that purpose; you just might not want to give it that high a percentage of the total module marks. If you want the assessment to motivate and help pace the students' studies, and to give them feedback on their learning, but you have not got the time to mark weekly sets of questions then computer-based MCQ tests available on the course's website might be the ideal solution.

The major problem with MCQs and other forms of objective tests is the time it takes to write good questions. Good detailed advice on this, and the various possible ways of using them with computers can be found at the Computer-aided Assessment Centre website: http://www.caacentre.ac.uk

Another possible solution is to see if you can find a textbook, especially a US textbook, that covers your topic area which may have banks of MCQs already written that come with the book. It would also be worth contacting the relevant LTSN Subject Centre for your discipline to see if they are storing any banks of questions (for contact details see http://www.ltsn.ac.uk).

Strategic reduction

If reductions are made strategically, there should be no detrimental effect on student learning. Reduction could be in the number of assessments themselves and/or in the amount of time spent giving feedback.

<u>i. Reducing assessment</u>

Examples of this would include:

• Identify and reduce repetitive assessment. It is not clear that writing lab reports every week for example is necessary either for students to learn from the lab work or to develop report writing skills. Indeed some such assignments have become so repetitive that they serve few useful functions.

• Use more varied assessment methods. Essays may be appropriate for some educational goals but can be expensive in staff marking time. A mixture of fewer essays backed up by objective tests might ensure both development of writing skills and reasoning skills, and the desired broad coverage of the material.

• Distinguish between courses with different goals. Some courses have aims, which may be best assessed through short tests rather than long essays or reports. Accept that different courses should have unique assessment patterns rather than all courses being assessed with the same type and volume of assignments. Other assessment strategies worth considering which might be both more appropriate to certain learning outcomes and shorter than some traditional methods might include: a briefing paper, an article, a book review, an annotated bibliography, a draft research bid, an A-Z, an encyclopaedia entry, a response to a client's question or a poster.

• Review the need for extensive summative assessment in the first year where most assessment is effectively pass/fail (in the sense of decisions concerning progression to the second year, rather than marks counting towards the final degree) and where almost all students pass regardless of the assessment system used.

• Assess only selected pieces of student work from a greater portfolio of work they have produced, possibly because it was a course requirement to do so.

ii. Reducing time spent on feedback

Examples of this, many of which we have covered already, would include:

• Give general feedback to the whole class rather than individual feedback

• Use feedback forms

• Use students

• Use statement banks

• Use computer-based tests

• Give focused feedback on only one different criterion each time. For example, in a series of English literature essays, for one you might focus your detailed knowledge of the author and the social and historical context, and for the next essay you might focus your feedback on the use of critical theory, and so on.

• Use audiotape to comment on students' work. While reducing the time you spend this may actually increase rather than reduce the amount of feedback given. As you read, speak your comments into a tape recorder. Do not write any down, but just indicate by numbers on the text (1, 2, 3...) to which specific sections your comments refer. Students frequently say that they get far more information from taped comments, including the tone of one's voice, than they do from written comments, and they also do not have to try to cope with some of our illegible writing. To make this system easy to operate students can be required to hand in a blank tape with their assignment.

Six case studies

Selected laboratory reports in Civil Engineering

In this case study marking time has been reduced to five per cent of what it was (through use of sampling assignments for feedback and marking, marking only part of an assignment, peer-assessment, self-assessment, model answers, and new course requirement).

A civil engineering department had a slow turnaround on the marking of regular practical reports, and a sense that this work was poorly focused. Students didn't understand why they were putting so much time in. It was unclear what the marking was for. Staff reviewed the objectives of workshops on their lab-based courses and decided they were mainly interested in four things - analysing and interpreting data, written communication skills, experimental design and understanding key concepts. Students were told: 'We want a portfolio from you at the end of the year that has at least 20 reports from the 25 practicals, and if you don't submit 20 then you fail the course and you aren't allowed to sit the exam.'

Students were also told that four reports would be pulled out at random. One would be assessed for communication skills, one for data handling, one for experimental design and one for understanding engineering concepts. Students didn't know which would be assessed for what, so they had to pay attention to all four things every time. There was only one bit of marking and it was very focused marking so it was extremely quick. It wasn't only that it was four reports out of 20, it was in fact only about 20 per cent of the four out of 20, so the marking time went down to about five per cent of what it was previously. (Obviously it matters where you sample from if students were improving during the course.)

The reports were date-stamped in the office and then put into the portfolio, so students couldn't put anything in the portfolio that hadn't been submitted on time. And they couldn't change them afterwards. When the deadline was passed, the lecturer would give feedback on that lab, using model reports or peer-marking exercises or self-marking exercises or through general feedback comments having sampled some of the submitted reports. They used a variety of cheap methods for students to engage with the quality of their work and others' work and think about standards. But they didn't use lecturer marking.

In addition to this huge saving in tutor time they also found that it actually improved the quality of the students' work. When they had had to submit 20+ reports all to be marked they knew that the marks for each only counted for a very small amount and therefore were prepared to be somewhat slapdash in their approach. Now only four count (and they do not know which four) they each contribute a significant percentage of marks so they now have to make sure that each report is of high quality.

Computer-based tests in Chemistry

This case study highlights how computer-based tests can provide students with regular feedback and help students to pace their learning over a module

A large introduction to chemistry module had the problem that many students were only studying it because it was compulsory for a range of other courses and were students who had studied very little chemistry previously or had bad experiences in chemistry and would not have studied it from choice. A very big problem was the amount of maths in the course, which frightened many of these students. The response of many of these students was to do as little work as possible during the module until just before the exam which was far too late. The tutor knew all too well that what they needed was to work from the very beginning of the module, especially on the mathematical problems, and to hopefully get positive and reassuring feedback on that work to help build their confidence. Unfortunately, the only way this could be achieved would be through regular assessed work and the numbers were too large for this to be possible if it required tutor marking.

The solution has been to create simple multiple-choice tests using freely available (to UK higher education) and simple to use software called the Castle Toolkit (see http://www.le.ac.uk/cc/ltg/castle/tools/). The tests are put on the course's website each week for the students to do in their own time. There are no marks for these tests, no record is kept of who has taken them, and they can keep repeating them until they get them all right. So what incentive is there for the students to take these tests? Well they know from the beginning of the course that the end of module exam will include a section of MCQs and that half of these questions will have been selected from those questions used in the weekly tests. This seems to have been sufficient incentive to inspire most of the students to attempt the weekly tests and there has been a definite improvement in the end of module exam results.

Peer marking in Electrical Engineering

This case study highlights how tutor assessment can be significantly reduced through the use of self- and peer-marking of examinations using model answers, and just some tutor sampling.

Over 100 students in a third-year physics class assessed their own performance and a peer's performance in a mid-term examination.

At the first class after the examination, students were given model answers, commentaries and a marking-schedule by the lecturer. Each student was allocated an anonymous examination paper, and set of model answers.

They were then required to fill in the space on the marking-sheet, saying where the student had departed from the model answers and awarding a score for each section (on a scale provided). Students returned the papers and marking-sheets the following week and received their own examination script.

They then applied the same procedure to their own paper without knowing what marks their peer gave the paper. The self- and peer-generated marks were then compared. If the range was less than 10% the student was awarded the self-assessment mark. Otherwise the paper was re-marked by a member of staff. In order to discourage students colluding with each other to fix marks, other papers were sampled at random and marked by staff. Students liked the system, staff reported time-savings, even allowing for extra time spent on preparing model answers and organising the movement of papers.

Strategic reduction of summative assessment in Education

This case study involves an education course which has managed to greatly reduce its summative assessment by strategically focusing the assessment of specific learning outcomes, overlapping assessment to more than one course, and use of self- and peer-assessment to provide feedback on student progress.

An education course has changed from a very expensive assessment system to one of the leanest systems seen anywhere. In the old system, students produced volumes of paper which lecturers ploughed through, but there was almost no assurance of standards. They couldn't tell whether the students were competent or not. You could tell that students were busy but not that they were competent. Tasks were too big and complex, and feedback came too late.

They didn't have any exams - it was all coursework. They have now agreed to have an exam - except that they call exams 'time-constrained tasks'. The exam tasks simulate - or in some cases are identical to – coursework tasks. For example, in the exam they will be given class records of pupils' performance in English in the national curriculum across a year. The data consists of one side of A4 plus a page of comments about pupils, and there will be a questions like 'Comment on the adequacy of the record-keeping system of the teacher' and 'Comment on the adequacy of the teacher's plan to cope with individual differences within the class with reference to the national curriculum, what you know about the teaching of English, what you know about the record-system.'

Hence, the exam pulls in things that are on the competency list. Students would know to expect a question about record-keeping where they would have to look at some real records. They'd know there would be a question about lesson planning and that they would be given some actual lesson plans. The only way of preparing was to look at record-keeping systems and make sense of them, and look at lesson plans and make sense of them. The ground rule is that students cannot be confronted in the exam with a task that they haven't tackled in a formative way during the year.

During the year teachers are encouraged to use model answers, peer feedback, self-assessment and class discussion. Students are set up in learning teams to help each other prepare for the coursework tasks and give feedback. Either side of each assignment deadline, these learning teams meet with tutors. At other times they meet without tutors.

Actual marking is confined to four exams ('time-constrained tasks') to assess the entire year. They have agreement between the different course leaders about the range of tasks, so there aren't two lecturers covering, for instance, lesson planning. There is a grid of competencies that students are supposed to address, and they can see that they are being covered by the exam. Exam answers are quite short, so the external examiner can see on a couple of sides whether students can do these things or not.

Peer-teaching teams in Accountancy

This case study describes how learning improves through the establishment of peer-teaching teams in a system that requires no additional assessment or feedback time to be spent by the tutor.

A 1st Year accountancy course had been stripped down to lectures and an exam, and performance was terrible. Very high failure rates. Almost nobody got marks over seventy, while about 40% of students got below forty. The majority of students got between 30 and 50, which was very, very low.

The tutor responsible bravely showed the students the appalling results from the previous year and put it to them that they might try a really radical experiment. 'Would you like a system where we set you up in teams of four and give you the average mark for your team of four, because I think the same thing will happen as happened in this other context, i.e. exam results will go through the roof?' And amazingly the students voted in favour of it.

They introduced it and it worked. They set up students in learning teams of four and told them: 'You will study in teams of four, you will sit the exam and tests as individuals, but you will get the average mark of your team of four.' The exam board (or the local quality assurance) stipulated three conditions:(i) that the students agreed to it, (ii) that no individual was allowed to pass the course if their individual mark was a fail mark, i.e. no-one could pass the course on the average of the others, and (iii) that no individual was allowed to fail the course if their individual mark was a pass mark.

Students taught each other furiously to make sure the average was high. The average mark went up to something like 56 and about 25% of students got over 70%. Students were still allowed to resit, and the resit mark was allowed to count towards a recalculated group average if it was better, so the group tutored students through the resits. For the first time in living memory nobody failed the course.

Schemes like this work best when tutors use their time with groups by discussing process - How do you learn effectively in teams of four? - rather than content. What made the difference was changing the process through the social dynamics. You could use economical assessment methods but it doesn't support learning unless the process is in place. The biggest increase in marks came from the better students. This demonstrated that the act of teaching others had the most impact on performance. Many of these processes bring the bottom up but this one spread the marks. Nothing changed except the social processes associated with the course. You can afford cheap economic assessment methods provided the processes are right. When lecturers shift to economic resource-based methods with low class contact and cheap assessment systems, success is often determined by the social dynamics - the way people collaborate and talk to each other out of class, and things like social pressure and peer pressure.

Assessed on how well they assess in Humanities

This case study describes how staff marking could be significantly reduced while students not only write an essay as before but now gain greater insight into both the subject content and essay writing skills by seeing five other essays and making judgements about them.

On a first year Humanities course the coursework assessment was an essay but as student numbers rose this became increasingly time-consuming to mark. So while still requiring the students to write an essay, instead of staff marking it the students were allowed to bring their essay and the marking criteria (which had been given at the same time as the essay was set) to a mid-term class test. In the test they were given five essays of differing quality on the same topic as the essay they had written, and they were asked to mark and rank them using the marking criteria and their own essay as a guide. They were then marked on how close their ranking and justifications came to what staff considered to be the correct order, which took considerably less time than marking an essay. When this marking was returned it was accompanied by a written explanation of the staff's marking decisions about the five essays. Staff on this course report that the quality of subsequent essays written has improved considerably since this strategy was adopted.

Bibliography

i. Acknowledgements

This paper reproduces with permission and develops work from the following three publications:

Gibbs, G. (1992) *Teaching more students 4: Assessing more students.* Oxford: Oxford Centre for Staff and Learning Development.

Brown, S., Rust, C. & Gibbs, G. (1994) *Strategies for diversifying assessment in higher education.* Oxford : Oxford Centre for Staff and Learning Development.

Oxford Centre for Staff and Learning Development (1997) *Assessment under pressure: 14 innovative case studies* [online]. Oxford: OCSLD. Available from http://www.brookes.ac.uk/services/ocsd/2_learntch/aup14pr.html

ii. References

Brown, G. et al (1997) *Assessing student learning in higher education.* London: Routledge.

Gibbs, G. (1992) *Improving the quality of student learning.* Technical and Educational Services. Bristol.

Gibbs, G. (2000) Are the pedagogies of the disciplines really different?. In C. Rust (ed.) (2000), I*mproving student learning - 7 Improving student learning through the disciplines.* Oxford: Oxford Centre for Staff and Learning Development.

Gibbs, G., Habeshaw, S. & Habeshaw, T. (1992) *53 Interesting ways to teach large classes.* Bristol: Technical and Educational Services.

Gibbs, G., Habeshaw, S. & Habeshaw, T. (1993), *53 Interesting ways to assess your student.,* Bristol: Technical and Educational Services.

MacLeod, D. (2001) Cash we must have. *Guardian.* 27th February 2001.

Nelson, C. (1994) Critical thinking and collaborative learning. In K. Bosworth and S. Hamilton (eds.) *Collaborative learning and college teaching.* San Francisco: Jossey Bass.

O'Donovan, B., Price, M. & Rust, C. (2001) The student experience of criterion-referenced assessment through the introduction of a common criteria assessment grid. *Innovations in Education and Teaching International.* **38** (1).

Price, M., O'Donovan, B. & Rust, C. (2001) Strategies to develop students' understanding of assessment criteria and processes. In C. Rust (ed.) *Improving student learning - 8 Improving student learning strategically,* [in press] Oxford: Oxford Centre for Staff and Learning Development.

Price, M. & Rust, C. (1999) The experience of introducing a common criteria assessment grid across an academic department. *Quality in Higher Education.* **5** (2).

Ramsden, P. (1992) *Learning to teach in higher education.* London: Routledge.

Rowntree, D. (1977) *Assessing students – How shall we know them?.* London: Harper Row.

iii Further reading

Angelo, T.A. & Cross, K.P. (1993) *Classroom assessment techniques.* 4th ed. New York: Jossey-Bass.

Brown, S., Race, P. & Smith, B. (1996) *500 tips on assessment.* London: Kogan Page.

Nightingale, P. et al (1996) *Assessing learning in universities.* Sydney: University of New South Wales Press.

The Learning and Teaching Support Network Generic Centre

The Learning and Teaching Support Network (LTSN) is a network of 24 Subject Centres, based in higher education institutions throughout the UK, and a Generic Centre, based in York, offering generic information and expertise on learning and teaching issues that cross subject boundaries. It aims to promote high quality learning and teaching through the development and transfer of good practice in all subject disciplines, and to provide a 'one-stop shop' of learning and teaching resources for the HE community.

The Generic Centre, in partnership with other organisations, will broker information and knowledge to facilitate a more co-ordinated approach to enhancing learning and teaching. It will:
- Work with the Subject Centres to maximize the potential of the network;
- Work in partnership to identify and respond to key priorities within the HE community;
- Facilitate access to the development of information, expertise and resources to develop new understandings about learning and teaching.

The LTSN Generic Centre Assessment Series Guides for:
Senior Managers
Heads of Department
Lecturers
Students

Briefings:
Assessment issues arising from key skills
Assessment of portfolios
Key concepts: formative and summative, criterion and norm-referenced assessment
Assessing disabled students
Self, peer and group assessment
Plagiarism
Work-based learning
Assessment of large groups

Published by
Learning and Teaching Support Network (LTSN)

For more information, contact the Generic Centre at:
The Network Centre, Innovation Close, York Science Park, Heslington, York, YO10 5ZF
Tel: 01904 754555   Fax: 01904 754599
Email: gcenquiries@ltsn.ac.uk
www.ltsn.ac.uk/genericcentre