

Theris
3928.
(vol. 2)

**CHANGING ASSESSMENT IN HIGHER EDUCATION:
POLICY, PRACTICE AND PROFESSIONALISM**

COLIN HOLROYD

EXTRA VOLUME: APPENDICES

VOL II

**THE INSTITUTE OF EDUCATION
THE UNIVERSITY OF STIRLING**

MARCH 2003

~~06/03~~

LIST OF APPENDICES

APPENDIX I	University of Stirling EdD Assignment 1
APPENDIX II	University of Stirling EdD Assignment 4
APPENDIX III	Theory Guidelines
APPENDIX IV	Subject Report: Chemistry
APPENDIX V	Subject Report Philosophy
APPENDIX VI	Subject Report Medicine
APPENDIX VII	Subject Report Design
APPENDIX VIII	Summary Report for University of Glasgow
APPENDIX IX	University of Stirling EdD Assignment 3
APPENDIX X	Are assessors professional? (Published paper)
APPENDIX XI	A research seminar: tasks and responses
APPENDIX XII	Aspects of approval and anonymity: the 'rules of engagement'
APPENDIX XIII	Interview schedule

Note Appendices I, II and IX are assignments which themselves have appendices. To avoid any confusion, the appendices listed above are separated in this volume by red sheets.

UNIVERSITY OF STIRLING

INSTITUTE OF EDUCATION

DOCTOR OF EDUCATION (EdD) PROGRAMME

UNIT 1 ASSIGNMENT

Colin Holroyd

This paper explores, with reference to the relevant literature, the assumptions underlying my research questions.

1. **Background to, and rationale for, research focus**

1.1 The University of Glasgow is in the process of devising an assessment policy applicable to the institution as a whole. A first version was issued in June 1998 and this is currently out for consultation. I was a member of the group who produced this first version; the group is to be re-animated in February 1999 and it is likely that I will be asked to re-join the group. (This experience gives rise to strand 1 of the proposed research.)

From 1995 to 1998, as a member of the Teaching and Learning Service of Glasgow University, I provided a series of workshops and courses on the practice of assessment; these were provided centrally and attended by members of staff from a wide range of different departments. I also visited a range of departments, on an out-reach consultancy basis, to work with departmental staff on assessment issues raised by them. (These experiences give rise to strand 2 of the research.)

Both sets of activities, those relating to institutional policy and departmental practice, fascinate me. The relationship between the two is obscure. (This gives rise to strand 3 of the research.)

1.2 The above may explain the relevance of the proposed research to me, but it is hardly a justification of its wider relevance, interest and importance. This justification lies in the following factors.

(i) In October 1998 I was made an Honorary Research Fellow of the University; the hope was clearly expressed that I would choose to research in the area of assessment in higher education. "This is of great interest to us and of considerable potential value."

(ii) I sought the opinions of the Vice-Principal (Learning, Teaching and Quality) and of the Convenor of the Assessment Policy Group. They expressed general approval, promised assistance in gaining access to appropriate people and welcomed research attention to the area. "This could be very useful to us - this type of research is greatly needed."

(iii) There seems widespread acceptance that assessment is gaining even greater importance in higher education. There are a number of reasons for this: expansion of student numbers and increased diversity in the needs and abilities of the student population put a strain on traditional assessment practices; students who have an increasing personal stake in financing their higher education exert pressure for greater transparency and fairness; the general public is increasingly interested in inter- and

intra-institutional standards; accountability and quality assurance procedures include the procedures and outcomes of assessment within criteria for judging quality; there has been a growth in understanding of the impact of assessment on the nature and depth of student learning; assessment is increasingly seen by course planners and curriculum developers as a central rather than an 'add-on' feature. Despite this growing importance, the amount of research on assessment policy and practice is relatively limited.

Although I am firmly convinced of the importance of assessment policy and practice as a research focus, I am a little apprehensive because of its scope. Because assessment is now seen as related to almost everything else, it is particularly difficult to keep research in the area focused and manageable.

2. The proposed research: aim and research questions

2.1 The aim of the research is to illuminate areas of assessment policy and practice in one Scottish university in the hope of making a modest contribution (a) to our knowledge of **how** assessment policy and assessment practice are changing, (b) to our understanding of **why** they are changing in these ways and (c) to our understanding of how changes in policy and practice are related.

2.2 Research questions: short form

RQ 1. How and why is the institutional policy on assessment changing?

RQ 2. How and why is departmental assessment practice changing?

RQ 3. What relationships exist between institutional policy and departmental practice - and why?

These three questions suggest three strands to the research; these strands are related to the three paragraphs of 1.1.

2.2 Research questions: expanded form

Research Questions

Some Links to Literature and Concepts

- | | | |
|----------|---|---|
| 1 (i) | How was the 1998 draft policy generated? | Nature of policy and strategy - control or influence; UK and Scotland policy context; assessment purposes; organisational image. |
| 1 (ii) | How do departments respond to this policy? | |
| 1 (iii) | Why do departments respond differently? | Disciplinary differences; departmental culture; policy acceptance or reconstruction. |
| 1 (iv) | How is the policy modified in light of responses? | |
| 1 (v) | Why is policy modified in these ways? | As in 1(i); central control vs local autonomy; uniformity and diversity. |
| 2 (i) | How is assessment practice changing? | 'Models' of assessment change. |
| 2 (ii) | What factors influence these changes? | |
| 2 (iii) | Why is assessment practice changing differently in different departments? | Disciplinary differences; departmental culture; educational ideology; student numbers and nature; technology; assessment theory; change agents etc. |
| 2 (iv) | What assessment issues are seen as salient? | Issues in assessment |
| 2 (v) | Why do issues have different saliency in different departments? | As in 2 (iii) |
| 2 (vi) | What further changes are wanted? | 'Models' of assessment change |
| 2 (vii) | What support is provided and sought? | |
| 2 (viii) | If there is a mismatch in 2 (vii) - why? | Conceptions of professionalism. |
| 3 (i) | What relationships exist between 1 and 2 and why? | Different conceptions of links between policy and professional practice. |

3. The proposed research: one project or two?

Acutely aware that I am working for two 'masters' and that there are differences in their expectations and requirements, I have decided that I should think of two research projects, A and B. Project A is for the University of Stirling and will extend from 1999 to 2002; project B is for the University of Glasgow and will extend from 1999 to 2004. Project B contains everything in project A and a bit more. A summary description of these two projects is given in Appendix I. The remainder of this paper is concerned with Project A only.

4. The research: an activity framework

Figure 1 attempts to show in diagrammatic form how different aspects of the work of the research (for Project A - Stirling) are related.

There are three foci of research activity: institutional policy is changing; departmental practice is changing; institutional policy and departmental practice are in some way related. Each focus has one main research question associated with it: RQ1, RQ2 and RQ3. Each research question has a 'how' component and a 'why' component. The 'how' questions will result in descriptive claims (about the nature of policy change, practice change and the relationship between the two). The 'why' questions will result in explanatory claims about these three aspects. It is suggested that the three sets of explanatory claims are likely to be related as shown.

Parts of the diagram are so cryptic as to be unintelligible. For example, the upper oval for RQ3 suggests a possible hypothesis about stance adopted on assessment issues. This is expanded in Appendix II.

Three rectangular boxes indicate some of the literature which will be relevant; literature about methodological concerns and the general context of the whole exercise are not included. The literature makes some claims more likely than others and then, later, should inform discussion of the issues.

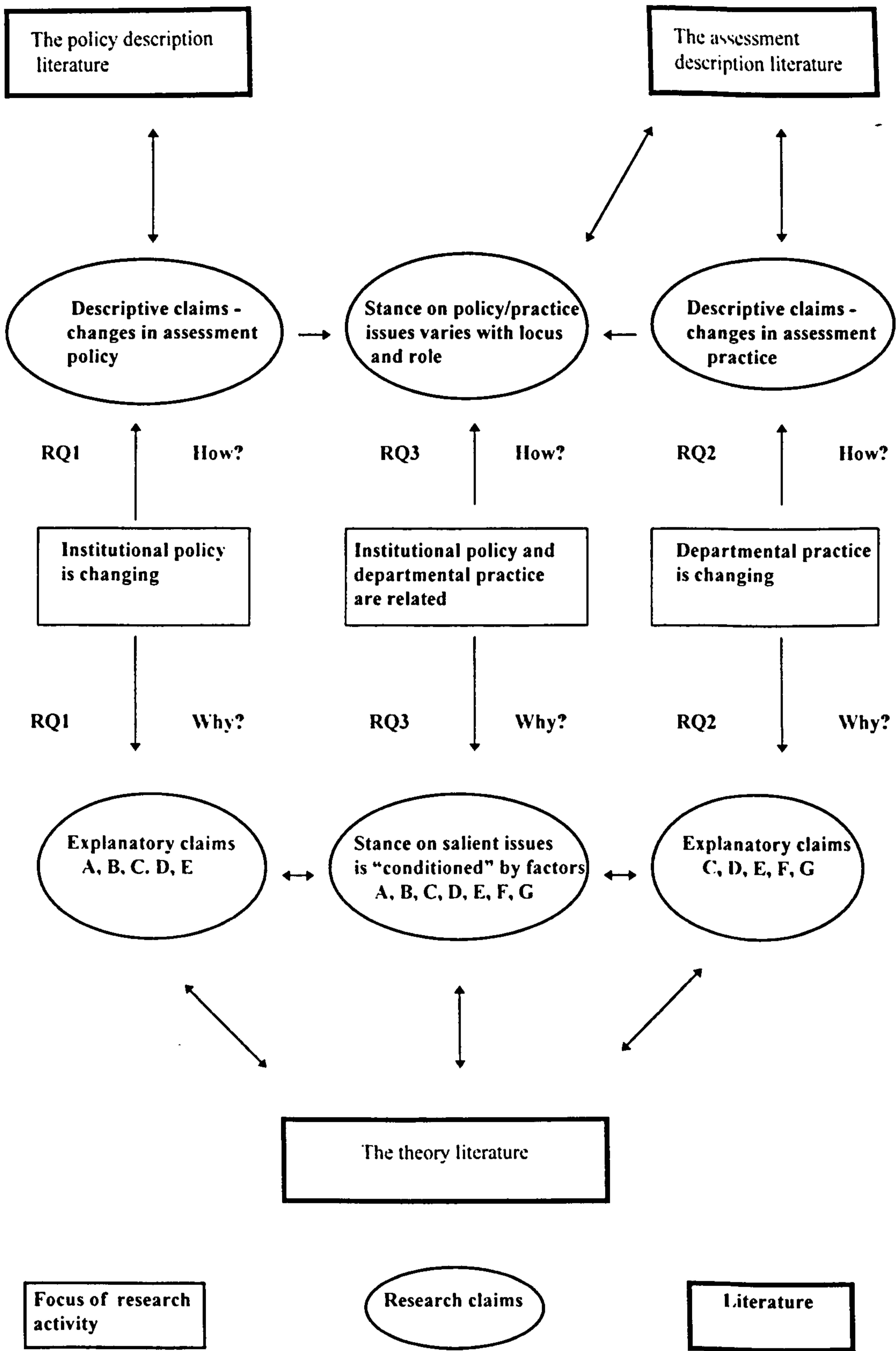
There are features of this diagram deserving comment. Firstly, it must be on probation; it has to be improved as time goes on. Secondly, it poses disconcerting questions. For example, why can something which occurs as a descriptive claim at the top left (a change in policy) re-appear as an explanatory claim at bottom right (a reason for change in practice)? Thirdly, this is suggested as an activity framework; it is not yet a conceptual framework. The latter will gradually develop after further immersion in the activity, the literature and the bath.

5. Methodological assumptions underlying the research questions

I attach one particular meaning to the word 'assumption': an assumption is a proposition accepted as true, without proof or evidence at this time, for the sake of future argument or action.

5.1 General methodological assumptions

a. The research required for the Stirling EdD must not be a-theoretical; it must go beyond mere data-gathering and make a contribution to understanding.



RQ = Research Question

Figure 1: Assessment policy and practice: an activity framework

The attempt to understand is necessarily nourished by theory. The term 'theory' is used here in its strong sense of an ordered relationship of concepts with explanatory import. The understanding of any complex reality arises from an attempt not just to describe it but to explain it (and scrutinise alternative explanations).

b. The research will lead to both descriptive and explanatory claims.

The distinction between descriptive and explanatory claims is a useful one, but it is not as easy as first appears. What, for example is 'theoretical description' and is Hammersley (1998) right to criticise it? It will be remembered that claims of these two types will have to be subjected to different procedures for assessing their validity.

c. The research will make no evaluative claims.

There will be no judgements like "In the light of criteria A, B and C this policy is unsound" or "This aspect of assessment practice is bad". The descriptive findings may, however, be used to inform evaluation by others. The complete avoidance of evaluation by the researcher should please some readers and have a positive effect in obtaining approval for both access and publication.

d. The research will not result in any prescription.

This will displease some readers who already expect positive recommendations for future action. It will be possible, however, to discuss the possible implications of claims and conclusions.

e. The research will distinguish between claims (arising from the actual case or cases studied) and conclusions (about the general research focus).

Claims about institutional policy will refer to one institution only and no general conclusions about the higher education sector will be drawn. If any reader can generalise as a result of their experience elsewhere, well and good. Claims from the case studies of departments can not be generalised to other departments by any form of empirical or probabilistic generalisation; however, some (appropriately cautious) theoretical inferencing or naturalistic generalisation will occur.

f. Issues will be important before, during and after the empirical work of research.

I have been involved in bits of research in which research questions were formulated, research activity was carried out, data were analysed and the report was almost written-up before the word issue was ever mentioned. "Right team - last chapter. What are the big issues? No more than seven..." This might be appropriate if the only issues to interest the researcher were those identified by the researched; issues would then genuinely have emerged from the data and it would have been premature to attempt to define issues before analysis of data started.

However, in most research issues are taken in by the researchers. This is not to say that the only important issues are those of the researchers (the etic ones); these will be supplemented by those which later emerge as crucial to the actors in the research (the emic issues).

I confess to having some difficulty in understanding the nature of an issue. I can describe it as something contentious and problematic, an important matter on which people take different stances, a problem about which people argue. But I do not know how to locate issues in relation to research questions and to theory. This is not a problem for Stake (1995): "I choose to use issues as conceptual structure... to force attention to complexity and contextuality" and "Issue questions provide a powerful

conceptual structure....” More and clearer thinking is required of me; at present I just have a strong feeling that issues should be central at all stages of the research.

5.2 Assumptions about relevant sources of data

a. Data sources appropriate to answering RQ1 (as defined in 1.1) are:

notes from meetings of the original policy group
working papers produced by and for the group
the first report of the group
responses from those consulted on this report
meetings of the re-constituted group (participant observation)
the members of the group (informal conversations and
interviews)
the final report.

b. Data sources appropriate to answering RQ 2 are:

people within the selected departments - interviews
the ‘ground level actors’ (the practitioners - those who
do the work of marking and assessment)
possible sub-groups would be - experienced
lecturers, recently appointed lecturers and
graduate teaching assistants

those responsible for departmental policy/strategy (head
of department, the ‘quality’ co-ordinator)

departmental and course documentation

key figures from the faculty and institutional context
dean of faculty or associate dean (education)
educational developers
senior managers

faculty and institutional documentation.

c. Data sources appropriate to answering RQ 3 are:

all sources listed in both a and b above.

d. Access to all the above data sources can be successfully negotiated.

The early signs are promising.

e. Students will not be a source of data in this research.

It could be argued that student views are desirable; certainly I would like to carry out research which recognised the importance of the student voice in this topic. Students have been excluded here, only to make the research manageable in scope. It is, however, envisaged that the assessment practitioners will be asked about their perceptions of student response to their practice.

5.3 Assumptions about methods of data collection

- a. A case study approach is appropriate to tackling the research questions.

For RQ 1 the 'case' to be studied is the generation of an institutional assessment policy. For RQ 2, there should be a small number (perhaps four) sites of departmental case studies. However, each of these will need to be considered in its faculty context (different for each department) and its institutional context. There is a little difficulty here I cannot at present see my way around: case studies are conventionally said to explore bounded systems and thus great care is needed in the definition of the boundaries. I may have to define boundaries, but I also want to recognise their permeability.

- b. The case studies envisaged are both instrumental and intrinsic.

I use these adjectives as defined by Stake (1995). Into each study I will import issues which perplex me, but I am not making the assumption that the issues which have occurred to me in advance are the important ones for those involved. Thus the methods adopted must allow for the emergence of those issues significant for the participants.

- c. Case studies should be characterised by methodological diversity.

A range of methods and techniques will be deployed. (Verma and Mallick 1999 - chapter 6) The use of any particular method will be justified by appropriateness to purpose. The variety of methods should allow an element of methodological triangulation; this has the potential to increase the validity of research claims.

The two main methods to be used are documentary analysis and interviews (probably semi-structured). However, there will also be opportunities for participant observation and informal conversations; these will require the making of field notes and the keeping of a research diary. Some use of respondent/informant validation will be incorporated.

5.4 Assumptions about target audience

- a. The main report will be written for the EdD examiners and will adhere to the standard conventions for such reports.

- b. This same report will be appropriate for the small number of people within the University of Glasgow who have approved the research and have an expectation that it will be useful to them.

This assumption may be re-considered in the light of later decisions on anonymity and confidentiality. This audience is composed of senior managers and educational developers.

- c. There will be a need for a summary version of the report.

This will be given to all who have assisted in any way with the research. It may also be appropriate for it to have wider dissemination within Glasgow University. It may, for example, be interesting to assessment practitioners in departments other than those studied; it may also have some value in the assessment courses/workshops run for recently appointed staff.

d. The research may give rise to findings that are of some interest to a wider academic audience; two suggestions of possible papers for publication are indicated in Appendix I.

5.5 Assumptions about levels of impact

a. As with much research, the main impact will be on the researcher.

Any research worth the name involves a substantial investment of time; in that time the researcher will be re-reading familiar sources, reading new things and regularly wrestling with perplexity about ideas and activities. The researcher cannot avoid being changed in many ways; these could be cognitive, affective or related to skill development. It is hoped that the algebraic sum of the changes will be positive, but this cannot be guaranteed.

b. The research will have some impact on a range of people through informal contacts.

If the research is important to the researcher it is almost inevitable that it will be the substance of informal interactions with fellow researchers, other colleagues, acquaintances and friends. Any one impact is likely to be very minor and the overall effect will remain unknowable; this does not mean it is negligible. Two quick examples of the effect on me of informal comments may help to illustrate the point. Firstly, in a recent meeting of our small group of EdD students, one member said something about his research which altered significantly how I think about my own. Secondly, on my first chat over coffee with the present director of the Teaching and Learning Service at Glasgow University, I asked about her research; she replied that her consuming passion was for research which was 'transformative, transgressive and emancipatory'; this phrase haunts me.

c. The research will have **some** impact on assessment policy and practice within the University of Glasgow.

Reports of one kind or another will go to a considerable number of people. It would be foolish to imagine that these will have a very significant effect on what people actually do. The relationship between research and any kind of educational practice is rarely direct. It is my hope, however, that the research will make some contribution to people's understanding of assessment issues; the way they think about assessment may be changed a little. There will be some probably weak and indirect impact on the decisions they make.

d. There may be outcomes worth disseminating more widely.

I have no ambition whatsoever to publish any old thing, just for the sake of it. However, if there was something which appeared worthwhile and to deserve writing-up, then I would certainly do it. There is so much published material in educational research and yet it seems, at times, to make very little impact at all. The impact is likely to be greater if any new research writing can develop existing ideas or build fruitful links between seemingly disparate sources.

I have once in the past been involved in some research which had a significant impact on the wider education community. On reflection, the reasons for the level of impact were (a) the research was in a policy-related area of concern to a range of key policy makers, (b) my co-researcher already possessed a national and international reputation, (c) aspects of the findings were published in a number of respectable outlets (writing these papers was only possible because I was unemployed for a few

months at the end of the project) and (d) dissemination was not confined to the written word - I gave thirteen oral presentations to various groups.

5.6 Assumptions about the ethics of enquiry

(a) Relevant codes of practice should be observed.

As a past member of SCRE and a current member of SERA I would wish to abide by their published codes. (SCRE 1996) (SERA 1997)

(b) Ethics are not merely an initial but a continuing concern.

It angers me when people remember ethical aspects during negotiation of access to research 'subjects' and then conveniently forget them. The ethics of the research process and of reporting are just as important. The principle of informed consent is of continuing relevance.

I am aware that explicating the purpose of the research is an ethical imperative. I am also aware that how the purpose is explicated can impact on the validity of later claims. It will be sensible to bear this in mind, but it is not likely to be a major difficulty in the present research.

(c) All reports will be written in such a way that individual sources are not named unless this is expressly agreed with the individuals concerned.

It would be possible to 'anonymise' the institution in reports, but it would remain easily identifiable to anyone who wished so to identify it. It is hoped that permission would be received from the Principal to name the researched institution as the University of Glasgow. The position with regard to the departments is rather problematic. It may well be that there are interesting differences related to subject disciplines; it would then be impossible not to name the departments. If permission for naming was not received, then the thesis for Stirling would have to be treated as confidential (for the limited period allowed) and there would be an unavoidable constraint on any wider dissemination. Whatever the eventual position on these two matters, all assessment practitioners (the 'ground-level actors') would have anonymity strictly respected.

(d) People own the facts of their own lives.

It seems to me essential to remember the warning from Walker (1978) that educational research can be theft. On the other hand, one cannot steal from someone what they freely give (and continue to agree they have freely given).

(e) There are ethical obligations to more than the 'subjects'.

The people researched are central in ethical considerations. It is also necessary to consider ethics in relation to customers, to colleagues and to the community. The treatment by Dockrell (1988) of these issues remains valuable.

5.7 Assumptions about researcher impact

(a) All research has an impact on what is being researched.

I cannot deal with this adequately in the space available. I will simply indicate the chapter headings I would adopt for a full treatment.

This assumption holds for research in physical science (Heisenberg).

The assumption is even more 'true' for social science and educational research.

Is there some objective external reality waiting 'out there' to be researched and which is not influenced by the researcher? (No)

How 'subjects' interpret the research situation is as important as how the researcher interprets it.

Is all research then a construction? (Yes - but we must still make reasonable assessments of whether any construction is more or less sound.)

(b) Interviews are never vehicles for pure information transfer.

The research proposed here makes heavy use of the interview method. There still exists a naive view that in any interview the interviewer asks a question and the interviewee simply responds by releasing the relevant information. It is necessary to accept that what any interviewee says is strongly influenced by his or her perception of who the interviewer is and whether or not the 'real' purposes of the interview are as the interviewer says they are. In other words the researcher impacts on the data provided in the interview. (Cohen and Manion 1994 - chapter 13)

The interview is an encounter which shares many features of everyday life; these features should be regarded not as obstacles to be removed, but as unavoidable features to be understood and taken into account. Cicourel (1964) lists five such features.

- (i) Many factors inevitably differ from one interview to the next - mutual trust, social distance, and interviewer control.**
- (ii) Any respondent may feel uneasy and adopt avoidance tactics.**
- (iii) Both interviewer and interviewee are bound to hold back part of what it is in their power to state.**
- (iv) Many of the meanings which are clear to one will be opaque to the other, even when the intention is genuine communication.**
- (v) It is impossible, just as in everyday life, to bring every aspect of the encounter within rational control.**

(c) The impact of the researcher is influenced by the role they are thought to possess (or to be playing).

One quick example: suppose I introduce myself as a researcher, associated with TLS, who happens to be a member of the Assessment Policy Group. Perception of my role by staff in departments then depends strongly on which part of that sentence gains most attention. The stereotypical responses might be: no researcher ever knows enough about any particular situation - one always has to waste time educating him/her about the special features of our discipline/context; staff from TLS are educational theorists trying to persuade us to adopt the latest fashionable practice; this is a spy from the Centre trying to uncover where we deviate from the party line.

6.1. General substantive assumptions underlying the research questions

- (a) A range of factors influence the generation, acceptance and modification of institutional assessment policy.
- (b) A range of factors influence change in departmental assessment practice; institutional assessment policy is only one of these.
- (c) There is a considerable overlap between the factors in (a) and those in (b) above.

6.2 Assumptions about the relevant literature

- a. In addition to the methodological literature, it will be necessary to develop understanding of the literature in four main substantive areas.

These are:

- the general context - higher education in Scotland and the UK
- the policy (description) literature
- the assessment (description) literature
- the theory literature (explanation)

The next four sections summarise how I propose to develop my knowledge and understanding; they should be seen as indicating assumptions about how best to enter the literature. They are assumptions **about** the literature rather than insights imported from it.

- b. Understanding of the general context requires that I learn more about how higher education in Britain has changed, continues to change and is predicted to change in the future.

I will continue my reading from four useful starting points. The first is the opening chapter (The Changing University? A Sketchmap with Coda) in *The Changing University?* (Schuller ed. 1995). The diagram on page 2 of this provides a useful organising framework for appreciation of the contours of change. The second, to deepen my understanding of how the changing purposes of higher education may link to assessment concerns, is the work of Barnett. (1994, 1997) The third is *Scottish Education* (Bryce and Humes eds. - forthcoming); the relevant chapters are 72, 73, 74 and 75. The fourth is regular scanning of the contents of The Times Higher Education Supplement.

- c. My understanding of the policy literature needs to be developed.

I am concerned here with those aspects of higher education policy which impact, directly or indirectly, on the assessment of students. My starting points are two: the appropriate sections of the Dearing (1997) Report and bulletins from the Quality Assurance Agency for Higher Education - particularly No 4 of October 1998. (QAAHE, 1998)

- d. My understanding of the descriptive literature of assessment needs to be updated.

There is an extensive literature here; I am tolerably familiar with much of it. I shall get back into the area by re-reading *Assessment Issues in Higher Education* (Atkins et al 1992) On order, but not yet received, is *Assessment Matters in Higher education*. (Brown and Glasner 1998) There is one book that has both substantive

and methodological relevance: *Researching into Assessment and Evaluation in Colleges and Universities*. (Ashcroft and Palacio 1996)

e. The developing conceptual framework will require considerable reading of the theory literature.

My conceptual framework is currently at an embryonic stage; it will keep changing and developing. Study of the theory literature will influence both the structure of the framework and later discussion of issues, claims and conclusions. I am mindful of the fact that assessing the validity of explanatory claims demands adequate knowledge of the possible alternative explanations.

Reading in the following areas will be required. One author only is indicated for each area.

Academics respond to pressures for change in a variety of ways. Important here is the extent to which policy is perceived as determinative or something which can be re-constructed.

(Trowler 1998)

The stance of individuals and departments on assessment issues is likely to be influenced by the nature of the discipline involved ("epistemological essentialism"), by educational ideology (perhaps traditionalism, progressivism, enterprise and social re-constructionism), by departmental culture and the ways in which 'great culture' impacts upon it (for example, to what extent are social inclusion, gender and anti-racism seen as vital concerns within assessment?). The huge range of references in the Trowler work gives useful prompts to the literature in all of these areas.

Stance on assessment issues will be related to how the organisation as a whole is perceived. Is the organisational image one of a community of scholars, a professional organisation, a political micro-system, or a bureaucracy? Is the organisation seen as adaptive and entrepreneurial or cybernetic?

(Middlehurst 1993)

Closely related to the above is how learning is conceptualised within a learning society (or organisation). People's stance on assessment issues will be related to what they see as an appropriate curriculum for the future.

(Young 1998)

Several points in the above relate to how academics see themselves as professionals. Do the respondents recognise teaching/ learning as an important area of professional expertise in its own right? Do they then see assessment as integral to teaching/learning?

(Nixon 1996)

Some informants are likely to put a strong emphasis on very practical reasons for changes in assessment practice. These are likely to include changes in the number and nature of students, assessment technology and general dissatisfaction with present activity. I do not know the recent literature about how these impact on assessment practice. My ignorance will be tackled via electronic search of the relevant journals and texts.

REFERENCES

- Ashcroft, K. and Palacio, D. (1996) *Researching into Assessment and Evaluation in Colleges and Universities*. London: Kogan Page
- Atkins, M., Beattie, J. and Dockrell, B. (1992) *Assessment Issues in Higher Education*. Newcastle: School of Education, University of Newcastle
- Barnett, R. (1994) *The Limits of Competence: Knowledge, Higher Education and Society*. Buckingham: SRHE and Open University Press
- Barnett, R. (1997) *Higher Education: A Critical Business*. Buckingham: SRHE and Open University Press
- Brown, S. and Glasner, A. (eds.) (1998) *Assessment Matters in Higher Education*. Buckingham: SRHE and Open University Press
- Bryce, T.G.K. and Humes, W.M. (eds.) (forthcoming 1999) *Scottish Education*. Edinburgh: Edinburgh University Press
- Cicourel, A.V. (1964) *Method and Measurement in Sociology*. New York: The Free Press
- Cohen, L. and Manion, L. (1994) *Research Methods in Education* (4th edition). London: Routledge
- Dearing, R. (1997) *Higher Education in the Learning Society*. Norwich: HMSO
- Dockrell, W.B. (1988) 'Ethical considerations in research' in Keeves, J.P. (ed) *Educational Research, Methodology and Measurement: an international handbook*. Oxford: Pergamon
- Hammersley, M. (1998) 'The Profession of a Methodological Purist' in *Doing Research about Education* ed. Walford, G. London: Falmer
- Middlehurst, R. (1993) *Leading Academics*. Buckingham: SRHE and Open University Press
- Nixon, J. (1996) 'Professional Identity and the Restructuring of Higher Education' *Studies in Higher Education* Vol 21, No 1, 5-16
- QAAHE (1998) *Higher Quality*. Bulletin No 4, October 1998 Gloucester: Quality Assurance Agency for Higher Education
- Schuller, T. (ed.) (1995) *The Changing University?* Buckingham: SRHE and Open University Press
- SCRE (1996) *Standards of Practice for Research* (unpublished) Edinburgh: Scottish Council for Research in Education
- SERA (1997) *Code of Practice for Educational Research in Scotland*. Edinburgh: Scottish Council for Research in Education
- Stake, R.E. (1995) *The Art of Case Study Research*. London: Sage

Trowler, P.R. (1998) *Academics Responding to Change: New Higher Education Frameworks and Academic Cultures.* Buckingham: SRHE and Open University Press

Verma, G.K. and Mallick, K. (1999) *Researching Education: Perspectives and Techniques.* London: Falmer

Walker, R. (1978) 'The conduct of educational case studies: ethics, theory and procedures' in Dockrell, B, and Hamilton, D. (eds) *Rethinking Educational Research.* London: Hodder and Stoughton

Young, M.F.D. (1998) *The Curriculum of the Future: from the 'new sociology of education' to a critical theory of learning.* London: Falmer

APPENDIX 1

PROJECT A : for University of Stirling

1. Changes in assessment policy in the University of Glasgow.
2. Case studies of change in assessment practice - four departments in their faculty and departmental contexts.
3. The relationships between 1 and 2.

Note The departments will be selected on the 'interestingness' criterion; there is no claim for typicality or representativeness.

By January 1999	reading (ongoing); development of proposal and conceptual framework.
By June 1999	documentary analysis; participation in policy group; proposal approved by SB, MW, RE; interview schedules for central people
By Dec 1999	interview of central people; analysis (ongoing); decisions on departments and types of people; access negotiations; case-study interview schedules
By June 2000	write up 'policy' section, possibly for publication; complete case study pilot and revise questions and procedures; case study 1
By Dec 2000	complete case studies 2 and 3
By June 2001	complete case study 4;
By Dec 2001	any return interviews required; design instruments for Project B; writing; secure funding for research assistant (Project B); plan Assessment Development Network (ADN)
By June 2002	complete writing-up; complete preparations for Project B (appoint research assistant and set-up ADN)

Main Outcomes of Project A

1. Thesis for Stirling
2. Questionnaire designed for institutional survey
3. Assessment Development Network set-up; ground level academics as action-researchers, possibly related to Cert THE.
4. Funding for Project B in place; research assistant appointed.

PROJECT B: for University of Glasgow

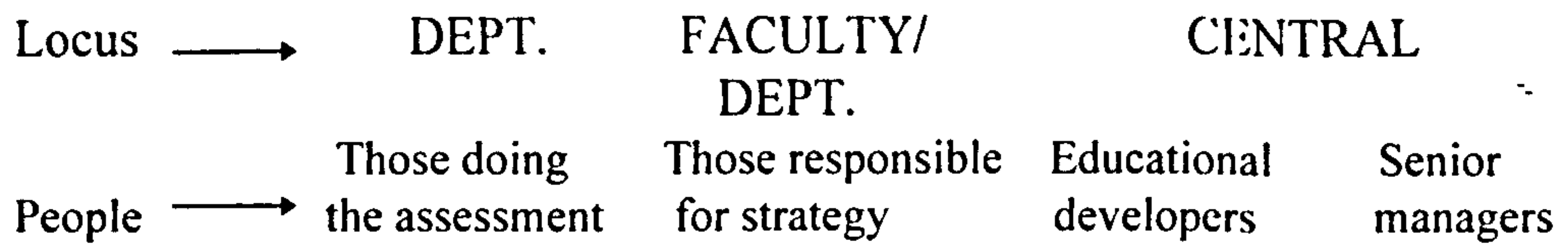
Changing assessment policy and practice in higher education: a case-study of one institution.

N.B. Project A contained case studies of four departments; project B claims to be a case study of the institution as a whole. The earlier case studies now inform an institution wide survey. In this project CH is no longer a sole researcher; there is a research assistant and a network of assessment developers (or action-researchers?) now in place.

By June 2002	Completion of Project A Questionnaire designed; funding in place; research assistant appointed, ADN set up
By Dec 2002	Questionnaire finalised, distributed and returned. (RA) Members of ADN keeping assessment dairies Paper for publication (case studies) (CH)
By July 2003	Survey analysed and written up ADN dairies completed
By Dec 2003	Full report for University of Glasgow: "Support for change" (?)
By June 2004	Complete the book.

The time scale is not totally insane. The Glasgow University policy for learning, teaching and assessment is to be implemented over a ten-year period - 1999-2009.

Hypothesis Stances taken on salient issues depend on institutional locus and role



Issues in Assessment

Certification/accountability v.s. learning enhancement	Tension	Certification/ accountability	Learning Enhancement	Certification/ accountability
Understanding v.s. competence	Tension	Tension	Understanding	Competence
Quality v.s. economy	Tension	Tension	Quality	Quality
Validity v.s. reliability	Validity	Reliability	Validity	Reliability
Threats to fairness	Student dishonesty	Student dishonesty	Assessor bias	Student dishonesty
What standards matter?	Personal and communicating group	Inter- departmental	Personal and communicating group	Inter- departmental
Others				

Why do stances differ?

- educational ideologies
- disciplinary factors
- cultural "traffic"

UNIVERSITY OF STIRLING

Edd PROGRAMME

UNIT 4 ASSIGNMENT

CHANGES IN THE ASSESSMENT OF STUDENTS IN HIGHER EDUCATION

A Research Proposal

Colin Holroyd

June 2000

Changes in the assessment of students on higher education

A research proposal

Preamble

There are conventional views that the research contained in a proposal will only occur after the proposal has been accepted and that the research actually carried out will be exactly as proposed. These conventions may be useful but they are difficult to harmonise with the conception of research design as a developmental and iterative process.

This proposal is for research activity which may be carried out from September 2000. It is informed by some pilot work which has already been done in April and May 2000. If the research goes ahead it may be different in minor aspects from what can currently be anticipated: there will be an element of progressive focusing and refinement between the work in subject area sites. The pilot interviews suggest that the form of the interviews is broadly acceptable, but the schedule is in need of revision (see Appendices E and F).

The research proposal is in five sections.

1. The Research Context
2. Purposes of the Research
3. Sites and Methods
4. Timescale
5. Analysis

References will be found on pages 17 and 18.

There are seven Appendices.

- A. A paper emerging from Unit 3 of the course, to be published in June/July 2000
- B. Choosing subject area sites
- C. A protocol - aspects of approval and anonymity
- D. Points from informal conversations
- E. First draft of interview schedule
- F. Thoughts from pilot interviews
- G. Musings on analysis

There is one very obvious omission from the assignment and that is the acknowledgements. The writer is grateful to a large number of people (including the research supervisor) who have given generously of their time and help. All the deficiencies in what follows are his alone.

CHANGES IN THE ASSESSMENT OF STUDENTS IN HIGHER EDUCATION

A Research Proposal

If you want to change student learning then change the method of assessment.
(Brown et al, 1997)

If we want to discover the truth about an educational system, we must look into its assessment procedures.
(Rowntree, 1987)

1. THE RESEARCH CONTEXT

The context of the proposed research can be described through consideration of five assertions: assessment is of great and increasing importance; assessment practice is changing; assessment policy is developing; the practice of assessment requires professionalism and assessment policy impacts on professionalism.

1.1 Assessment is of great and increasing importance

Three broad purposes of assessment in higher education can be distinguished: the first relates to the certification of students, the second to the accountability of institutions for the courses they provide and the third to the enhancement of learning. In the first case assessment provides the evidence on which judgements are made about the eligibility, or otherwise, of students to receive certain awards which signal completion of a stage of higher education or fitness to practise or both. In the second case assessment is one important source of evidence that the institution is in fact doing what substantial sums of public money are given to it to do. In the third case, recognition of the power of assessment to influence the quantity and quality of student learning exerts an obligation for assessment to be used to positive effect.

These purposes taken together mean that 'the importance of assessment can scarcely be over-emphasised'. (Freeman and Lewis, 1998 - p 7) Assessment is important

'..... to students, the tutors who assess them, the institutions in which they are assessed, the parents, partners and carers who support them; it matters to the employers who would like to offer them jobs on graduation and to the funders who pay for higher education and want to see the maintenance of standards and value for money.'

(Brown and Glasner, 1999 - Preface)

The importance is great; it is also increasing. The factors which influence this are well-known. Student numbers have increased dramatically as has the diversity within the student population. As more money is invested in higher education and the government is increasingly concerned with

accountability, ever greater pressure is exerted on assessment procedures to perform their certification function effectively and transparently. At the same time, much more is understood about the nature of the backlash that assessment has on learning; assessment is now coming to be recognised as the single most important influence, for good or ill, on learning. Students take their cues from what is assessed rather than what is asserted by lecturers to be important. (See, for example: Boud, 1995 - p 39; Biggs, 1999 a - p 141; Cowan, 1998 - p 137; Erwin and Knight, 1995 - p 181; Miller et al 1998 - p 3 and Toohey, 1999 - p 170.)

If there is a wish, and there should be, to enhance student learning, then the methods of assessing that learning should be changed and improved. Hence the first quotation at the beginning of this proposal.

1.2 Assessment practice is changing

It is normal for texts on teaching, learning and assessment in higher education to give summary accounts of how assessment practices have changed over the last few decades. These accounts are not identical, but there is substantial overlap. To take just five recent examples, important features of change are identified by Biggs (1999 a - chapter 8), Brown et al (1997 - p 13), Brown and Glasner (1999 - chapter 2), Freeman and Lewis (1998 - p 310) and Toohey (1999 - chapter 9).

A summary conflation from these sources identifies these significant dimensions of change:

- (i) increasing emphasis on the learning enhancement purpose of assessment rather than its certification and accountability purposes;
- (ii) increased attention to formative rather than summative purposes;
- (iii) more emphasis on a standards model of assessment, involving criterion referencing, and less on a measurement model, involving norm-referencing;
- (iv) more frequent provision of descriptive comment and constructive feedback and less restriction of assessor response to marks, grade and summary labels;
- (v) a move from dependence on one main method of assessment (and end-of-course assessment) to deploying a variety of methods (and within-course assessment);
- (vi) less reliance on assessment by teaching staff alone and more involvement of self, peers and workplace assessors;
- (vii) increased emphasis on assessment as integral to teaching rather than as a separate activity occurring after teaching.

To identify these as system-wide trends must allow for local contravention of such trends; there are, no doubt, many places and courses where change has been in an opposite direction. There are also some pervasive countervailing pressures which could in the future result in significant reversals. For example, it is frequently claimed that validity is in the ascendant in assessment thinking and activity; this is evidenced by moves to more authentic forms of assessment and the insistence that assessment procedures be aligned more constructively with statements of intended learning outcomes. (Biggs, 1999 b) However, others would argue that pressures for accountability and consistency from government and for transparency from potentially litigious students-as-consumers mean that reliability is a more compelling requirement than ever before.

Relatively recent moves towards modularisation of higher education courses and programmes have

resulted in much reported change to assessment procedures. These changes are helpfully summarised and discussed by Wolf and her co-authors in the report *Assessment in Higher Education and The Role of Graduateness*. (HEQC, 1997)

1.3 Assessment policy is developing

In the past, policy on assessment in higher education tended to reside in departments and faculties. Although it would not be true that national policy has only started to emerge in the last 2 or 3 years, it is only since the report of the National Committee of Inquiry into Higher Education (NCIHE, 1997) - the Dearing Report - that national policy has come to be of high significance. What that report actually recommended to government on assessment has been described elsewhere. (Holroyd, 2000 - Appendix A) For present purposes, it will be sufficient to consider briefly two developments only. Recommendation 25 has resulted in the setting-up of small expert teams to provide benchmark standards/statements which define for the UK as a whole, and for the first time, what has to be assessed within honours degrees in forty-two different subject areas. Recommendation 24 gives the Quality Assurance agency (QAA) the remit of developing a code of practice, including sections on external examining and student assessment, to be adopted by all institutions as a condition of continued public funding.

The draft codes on external examining (QAA, 1999) and on student assessment (QAA, 2000) are written in the same general form. They articulate precepts with accompanying outline guidance. The precepts 'identify those key matters which the Agency expects institutions to be able to demonstrate it is addressing through its own quality assurance mechanisms'. The guidance is intended to be neither prescriptive nor exhaustive, but in many institutions 'the guidance will constitute good practice'. If a code is a set of rules or a body of laws, if a precept is a command or rule and if guidance defines what should 'normally' be seen as the good practice required for the continuation of funding, then it seems as if we do have a national policy which it will be difficult for any university to ignore.

The essence of the developing national policy is that institutions must develop an institutional policy in conformity with the national precepts. As an extension of this, it seems inevitable that institutions will require faculties and departments to generate and operate assessment policies which conform to the institutional and national policies. Traditionally assessment policy has been developed and then formalised at departmental and faculty levels with little or no central direction. The increase in policy making can be seen as policy in a vertical sense (downwards from nation to institution to faculty to department); this will be accompanied by much policy activity within and across institutions i.e. policy in a horizontal sense. This useful distinction between vertical policy-making and horizontal policy activity derives from Colebatch (1998). Assessment policy is thus rapidly developing into a contested terrain in university settings. (Ozga, 2000)

1.4 The practice of assessment requires professionalism

It is common (although not common enough) for academics to ask what their assessment procedures require of those assessed i.e. their students; it is rare to ask what assessment requires of those who do it i.e. the assessors themselves. One attempt to answer this neglected question has been essayed

by the present writer. (Holroyd, 2000) He concludes (i) that assessment does not have one specialised knowledge base, but necessarily draws on several, (ii) that decision-making in most, but not all, aspects of assessment is difficult, complex and discretionary, (iii) that assessment should be crucially concerned with justice for students (in its certification function) and with their empowerment (in its learning enhancement function) and (iv) that effective and fair assessment increasingly require that assessors form members of a regularly communicating group - it is becoming more and more important for assessors to co-operate, to collaborate and to communicate.

Traditional professionalism has been defined in a number of ways (see, for example, Freidson, 1994). These stress such things as the specialised knowledge and skills required, primary allegiance to some self-controlling body and the taking of difficult decisions for the client within some ethical concern for the well-being of society as a whole. If traditional professionalism has been discredited, then a new professionalism can be claimed on different grounds. For example, Nixon et al. (1997) see a regenerated professionalism as necessarily involving collegiality, negotiation, cooperation and partnership. This new professionalism will have to consist of a set of practices 'imbued with an ethics of integrative action that seeks to accommodate differing values and cultural outlooks'. Bottery (1998) covers similar ground but provides a rather different account of what these embedded ethics might be: he suggests an appreciation of the provisionality of all knowledge and a commitment to truth seeking, reflective integrity and humility, and client empowerment. These are the values needed to encourage the development of a strong democracy.

Assessment requires certain abilities and qualities in the assessors which are related to one or other of the above conceptions of professionalism. For present purposes, it matters not at all whether professionalism is defined in traditional or emerging terms. The practice of assessment requires professionalism.

1.5 Assessment policy impacts on professionalism

The developing national policy on assessment is regularly perceived as increasing the pressure towards homogeneity and uniformity; it restricts the right of institutions to be different, it constrains their autonomy. Developing institutional policy restricts the diversity across departments and faculties; it constrains the autonomy of individual faculties and departments. The developments in policy at all levels can similarly be seen as constraining the professionalism of individual academics.

While this argument has some validity, it seriously underestimates the complexity of the issues. A glimpse of the problematic nature of the links between policy and professionalism can be gained from the following.

- (i) The autonomy of individuals in assessment practice has always, rightly, been constrained and a change in the nature of the constraints is not self-evidently wrong.
- (ii) One explicit purpose of policy can be to increase the professionalism to be deployed within assessment practice. (For example, Recommendations 13 and 48 of Dearing are concerned with the training, accreditation and continuing professional development of university teachers.)
- (iii) The impact of policy on professionalism may not be as one snooker ball upon another, but rather as an outline script for a drama to be played out by professional actors in an agonistic space.

2. PURPOSES OF THE RESEARCH

2.1 Focus and aims

The research will focus on changes in the assessment of students in an institution of higher education. This focus has been chosen for two main reasons: firstly, student assessment is of great and increasing importance (the first assertion above) and secondly, the focus embraces the three main emphases in the Stirling EdD programme namely, the understanding of institutional change, the nature of academic professionalism and the impact of policy on practice. The reason for choosing one institution is dealt with in Section 3.

The research aims to illuminate and increase understanding of changes in assessment policy and practice; it aims to describe and explain, but not to evaluate. Four rather more specific aims can be explicated. This research aims to describe and explain, for this one university,

- (a) the changes in assessment that have occurred and that are anticipated and desired,
 - (b) the policy that has developed on how students are to be assessed,
 - (c) the relationships between the changing practice and the developing policy,
- and (d) the impact of changing practice and developing policy on the professionalism of the teachers of this university.

2.2 Research Questions

The empirical work of the research will be guided by a number of research questions. These questions have been refined since their initial formulation (EdD Assignment I) and may well undergo further alteration, given the developmental and iterative character of research design. Two further things are worth noting: (i) the research questions are stated not in order of importance, or as suggested by the aims, but as they relate to the sequence of research activities outlined later and (ii) they are given in the tenses presently appropriate, but will be translated into the past tense for the final report.

- (i) What does emerging assessment policy say about how students are to be assessed?
- (ii) What changes have there been in how students are assessed and why did these changes occur?
- (iii) What changes in assessment are likely to be required and why?
- (iv) What changes in assessment are desired and what conditions are necessary to make those feasible?
- (v) How have past changes affected, and how will anticipated changes affect, the work of academics?

The above research questions could well have been formulated before much thought was given to defining the research context or to planning the research approach. These activities have so far resulted in four further questions.

- (vi) To what extent do the changes that have occurred in this university conform to general patterns of change posited for higher education as a whole?
- (vii) Do the reasons given for change correspond to those posited by others?
- (viii) In what ways do the reasons given for past changes differ from those given for changes likely to be required in the future and why?
- (ix) Does the stance on assessment purposes in different subject areas agree with the assumptions implicit in the criteria for choosing those subject areas?

2.3 A note on boundaries

Before leaving the purposes of the proposed research it is necessary to clarify what is meant by 'how students are assessed'. In earlier versions of the research questions, the key term was 'methods of assessment'. It has now been decided that this is unhelpful; 'methods' is a term that is used confusingly in the assessment literature. In some texts 'methods' are synonymous with techniques or instruments; in others 'methods' are equated with broad approaches (such as formative and summative assessment); in yet others 'methods' cover the sources of assessment data (such as other people). In this research, the term 'methods' will be avoided wherever possible and 'how students are assessed' will initially be defined to cover the following:

- the techniques by which students are assessed;
- the timing within a course when assessment occurs (continuous or terminal assessment);
- whether or not assessment counts for a final grading or award (summative or formative);
- who conducts the assessment (tutors, self, peers, workplace assessors);
- the nature of feedback provided (descriptive comment, summary mark, grade, label).

Although this definition is worryingly broad (and may later have to be limited) it should be noted what it excludes. It is not intended that, within the empirical work, the researcher will introduce such topics as these:

- the design and management of assessment systems;
- the creation and testing of assessment items and instruments;
- level and grade descriptors;
- procedures for the aggregation of grades and final decision-making;
- the administration of assessment - invigilation, record-keeping and reporting;
- provision for students with special needs;
- sources of error - bias in assessors and dishonesty in students;

the selection and deployment of external examiners;
the selection, training, monitoring and professional development of internal examiners;
mechanisms for appeal against assessment decisions.

These exclusions may well be seen as important by informants and introduced into interviews. Such perceptions and introductions will be of interest. It may also be that discussion of issues arising from the research will require attention to some, or all, of the above.

2.4 No bigger purpose?

The research aims to cast a little light and to provide some answers to a few questions. This sounds appropriately modest, but some may think it trivial. Is there no bigger purpose?

This research seeks to make a contribution to the purposes which should be shared by all educational research. If the fundamental criteria for assessing research are relevance and credibility (Hammersley, 1998), this is because its fundamental purposes should be (in contemporary parlance) 'to make a difference' to the quality of learner experience and to enhance some theoretical framework in credible ways. It is hoped that this present research will have some positive impact, however small and indirect, on student learning in one university; it is also hoped that some inchoate fragments of theory can be clarified and more elegantly arranged. (This point is revisited in Section 5.)

The inclusion of the word 'contribution' in the above is meant to preempt criticism that the grander purpose is merely grandiose.

3. SITES AND METHODS

3.1 Choice of sites for research activity

It is proposed that the research be carried out in one institution of higher education, the University of Glasgow. (For convenience this will be referred to hereafter as the University.) Some thought was given to working in several Scottish universities and rather more thought to working in two contrasting universities (Glasgow and Stirling). The decision to limit the study to one institutional site was taken on grounds of manageability, practical convenience and obligation, with an awareness of the competing benefits of breadth and depth and of the implications of the decision for the generalisability of claims.

An early intention was to carry out an instrumental case-study (as this is defined by Stake, 1995) of changes in assessment policy and practice in the University. This was quickly recognised to be unrealistic in scope for one part-time researcher. However, there still remains a possibility that the research currently proposed could be nested within a larger enterprise which could legitimately claim to be a case-study of an institution. (EdD Assignment I - Projects A and B.)

The proposal now is for a more modest exercise which has two strands: the minor strand will focus on policy development and address research question (i); the major strand will focus on changing

practice and will address the remaining research questions.

The 'site' of the minor strand is the Assessment Working Group of the University. This group was set up in 1997 and its remit has been variously described as the production of a strategy, a policy, a code or a guide for assessment practice. The proposer has been a member of this group since its inception.

The sites for the major strand will be four departments within the University. Again the plan has been modified. At first it was to produce four departmental case-studies of change in assessment practice. It then seemed (a) unrealistic to cover all the assessment activity for all courses : undergraduate and postgraduate within a department and (b) unwise to describe as a 'case-study' (which one would expect to be characterised by a diversity of methods) research which would essentially rely on only two methods (documentary analysis and interviewing). Thus there will be four departmental sites, but the research should be thought of as an interview-based study of four subject areas at two undergraduate levels, Level I and Honours. The subject area studies will be sequential; although the core concerns will be the same in each case the approach will be slightly modified from one to the next; as the researcher learns, there will be some progressive focusing.

Four departmental sites have been identified. Why were these sites selected? There is a widely held view that approaches to assessment are strongly influenced by the nature of the subject assessed. (See, for example, Bridges et al (1999) and as yet unpublished research by Mason at St Andrews.) This would suggest that the subject sites should be selected from various points on a 'hard-soft' continuum; they might be mathematics, chemistry, sociology, English literature. This was seriously considered but it seemed inadequately to reflect the wide diversity of subject areas within a complex institution. It was thus decided to hypothesise that subjects varied in their assessment approaches along two orthogonal axes: the first captures the extent to which subjects are perceived as 'pure' or 'applied' and the second the relative emphasis placed on understanding a bundle of defined content as opposed to evaluating the quality of alternative arguments. Where any subject is perceived to stand in relation to these two axes can be inferred from answers to two relatively simple questions about assessment purposes at Honours stage. Are your assessments primarily concerned with enabling judgements about academic soundness or about fitness to practice? In your assessments how much weight do you attach to student understanding of key concepts and how much to ability to judge opposing arguments (or solutions or products)?

If this analysis holds, it seems sensible to choose subject areas from the outer corners of the four cells defined by these axes. The proposer conjectures that four appropriate subject areas are chemistry, philosophy, medicine and design. (Appendix B) Three points are worth noting. Firstly, this choice of sites has the added attraction of involving Glasgow School of Art as an associated college of the University. Secondly, although treating each of these as a single subject area may be thought problematic it is seen as defensible by the people responsible for them. (Chemistry used to be thought of as three subjects but is now considered one; Medicine is one subject area within the new medical curriculum of the University although there are over thirty contributing departments.) Thirdly, these criteria for the choice of sites have provided the rationale for research question (ix).

Within each subject area site, normally eight people will be interviewed. These will be the following, or their equivalents:

- the head of department (or a senior member of staff nominated by him/her)
- the chair of the teaching and learning committee
- the class coordinators for Level 1 and Honours
- the assessment/examination officer
- three members of staff with a significant involvement in assessment
 - one with long experience in the University
 - one with experience in another university
 - one 'junior' member (a graduate teaching assistant or an assistant or probationary lecturer).

As departmental practice occurs in a faculty and institutional context it may be necessary to interview people with a responsibility for assessment at such levels. This will be decided after the research activity in the subject sites.

3.2 Choice of research approach and methods

A somewhat eclectic approach to methods will be adopted; by this is meant that methods will be selected on a judgement of their fitness to the range of purposes outlined, bearing in mind that the fundamental requirement of methods is that their outcomes merit respect. There is not space in this proposal for a detailed justification of this stance. In outline, it depends on the belief that although theoretical perspective and underlying epistemology do proscribe certain approaches (or methodologies) and methods (or techniques) they do not prescribe with any exactness what should be chosen from the available range of approaches and methods. (Crotty, 1998 - chapter 1)

The methodology is essentially qualitative and phenomenological, based on an interpretivist perspective and underpinned by a constructionist epistemology. The main methods to be employed are semi-structured interviewing and documentary analysis; there will be a small element of participant observation (of the Assessment Working group). If theme identification, data reduction, content analysis and comparative analysis are admitted as methods (Crotty, 1998 - page 5) these will also necessarily be involved.

Within the interviews, the first and major part will aim to elicit the perspectives of the interviewees, their memories and their predictions and the meanings they attach to certain concepts. However, the second and minor part of each interview will try to test the extent to which views held (or posited) by others are shared by the informants (see research questions vi and vii); this part of the interview will be structured and could be described as the oral administration of a questionnaire. This element can be seen as positivist and will allow a little quantitative reporting (but no sophisticated statistical analysis).

At this point the second quotation chosen to open this proposal deserves comment. It is vitally important that we look into assessment procedures, but any notion that there is one truth about the educational system waiting there for some researcher to discover is chimerical given the predominant theoretical perspective of the research.

3.3 The preparatory phase

This proposal in its present state has been informed by some preliminary work of three types: informal conversation, negotiation of access to sites and piloting of a draft interview schedule. The plans for this phase and its outcomes have been discussed with the research supervisor.

3.3.1 Informal conversations

It was decided to hold informal conversations with seven people. The invitation to each was based on 'I have a very embryonic idea of some relatively small-scale research I might do. How do you react...?'. A 'conversation framework' was prepared and notes were taken on this during each conversation. The framework had four sections:

- (i) an outline of the proposed research;
- (ii) ways in which subject area sites might be chosen;
- (iii) people who might be interviewed in each site;
- (iv) diplomatic and ethical concerns surrounding approval and anonymity.

Everyone approached readily agreed to talk. The people (all within the University) were:

- (i) the Vice-Principal (Learning and Teaching);
- (ii) the Director of the Teaching and Learning Service (TLS);
- (iii) a member of TLS with research expertise;
- (iv) a member of TLS with assessment expertise;
- (v) a senior lecturer in psychology described as hostile to research in paradigms other than his own;
- (vi) an associate dean, well-versed in the ways of the University and convener of the Assessment Working Group;
- (vii) a former head of a large department.

The atmosphere of the conversations was informal in all cases; in the first two cases, however, there was one formal purpose, namely to secure approval in principle for the research to proceed. This approach was freely given. The Vice-Principal recommended the production of a written protocol on aspects of approval and anonymity. Such a protocol was drafted and appears as Appendix C; it was discussed in those conversations which took place after its production. Attention is drawn particularly to the careful wording of the final section. This is meant to recognise the rights of the participants and the obligations of the researcher.

A summary of the main points emerging from the seven conversations is given in Appendix D.

3.3.2 Negotiation of access to sites

Negotiations have been successfully completed with appropriate people in the four subject area sites to continue with the research. The people are:

- (i) the chair of the teaching committee of the department of chemistry (responsibility having been delegated to him from the head of that department);
- (ii) the head of the department of philosophy;
- (iii) the associate dean (education) of the faculty of medicine;
- (iv) the chair of design and craft.

Agreement has been secured to approach eight people in each area at various times from September 2000 with a view to their being interviewed. Recommendations as to who the people should be have been obtained. Individuals remain free to refuse.

The initial approach to each of the four people was similar. Thereafter, in one case the process was quick and easy and in a second it was protracted and difficult (procedurally, although personally delightful). These two processes can be summarised as follows.

- Case A
 - Informal phone call
 - Sending of formal letter and protocol
 - One exchange of e-mails
 - Meeting: approval immediately given without question; potential interviewees decided; e-mail from the approver to nominees - 'I am sure you will want to cooperate fully...'
 - Letter of thanks from proposer.

- Case B
 - Informal phone call (successful on fourth attempt)
 - Meeting: tentative move towards approval
 - Sending of formal letter and protocol
 - Four exchanges of e-mails
 - Request discussed at departmental meeting
 - Exchange of e-mails: 'the request was received with apathy from most and antagonism from a few'
 - Second meeting: approval gained, nominations made - 'expect some refusals....'
 - Letter of thanks from proposer

3.3.3 Piloting of interview schedule

Three pilot interviews took place. These had two purposes: to try out the first draft of the interview schedule (Appendix E) as a research instrument and to remind the proposer of his strengths and weaknesses as an interviewer.

Those interviewed were chosen to represent some of the categories of people to be encountered later. They will not have to be interviewed again. They were:

- (i) an experienced lecturer in an arts subject, known to have introduced group-work and oral presentations into the assessment of his subject;
- (ii) an associate lecturer in a science subject, known to have worked on computer-assisted assessment of her subject and to be heavily involved in

- Level 1 marking;
- (iii) a recently-retired senior lecturer in chemistry with many years of involvement in and responsibility for assessment at both Level 1 and Honours stage.

The interviews were conducted in a place chosen by the subjects and were recorded. They lasted between 35 and 65 minutes. The interviews were not transcribed. Each tape was thereafter played three times. On the first occasion the tape was played straight through to gain general familiarity with the data. On the second occasion, the tape was stopped and replayed very frequently; copious notes were taken (including evaluative comments on the interviewer) and checked. On the fourth occasion the tape was played through to regain an holistic impression, stopping only to transcribe verbatim quotations which might later be useful.

A summary of the salient points emerging from the pilot interviews appears in Appendix F.

4. TIMESCALE

It is proposed that the research be carried out according to the timeframe summarised below.

April - May 2000	Preparatory phase	informal conversations access negotiations interview piloting
June 2000	Write research proposal	
August - September 2000	Revise research design in the light of critical feedback Analysis of documentation and research notes from Assessment Working Group Arrange interviews in chemistry and philosophy Revise interview schedule	
October - November 2000	Interviews in chemistry and philosophy	
December 2000	First drafts of 'chapters' on chemistry and philosophy to interviewees Arrange interviews in medicine and design	
January - February 2001	Interviews in medicine and design	
March 2001	First drafts of 'chapters' on medicine and design to interviewees	
April 2001	Interviews at faculty and central management levels (if seen as necessary/desirable)	

May - June 2001	Next-to-final versions of four 'chapters'
September 2001	Updating of materials from Assessment Working Group Design structure of final report
October - December 2001	Write chapters on context, purposes, literature and methods
January - March 2002	Write final versions of 'findings' chapters
April - June 2002	Write chapters on claims and conclusions, issues, and theory development
August 2002	Organise appendices, write acknowledgements, compile references Decide whether the research can be extended to be of more value to the host institution.
September - October 2002	Final revision and submission of thesis

Notes on the above

- (i) Analysis is specifically mentioned (in relation to the minor strand) for August/September 2000. This is because otherwise it is likely to be forgotten. This inclusion should not, however, mislead. Analysis will be ongoing from August 2000 to August 2002.
- (ii) Sourcing and study of relevant literature will take place periodically throughout the whole period.
- (iii) Revision of the interview schedule is explicitly mentioned for August/September 2000. There is likely to be some refinement of the schedule (and indeed of the overall design) between work in subject area sites.
- (iv) The timetable is demanding from October 2000 until March 2001 and may be thought over-generous thereafter. There are reasons for this. In the practical activities of research the proposer works better to deadlines which are tight, but which permit rather than encourage a modest amount of slippage. In the more cerebral activity of writing the proposer recognises that he works neither as fast nor as consistently as (he thinks) he did in the past. The timetable should allow for some progressive mental deterioration.
- (v) It is hoped that at least one publishable academic paper on some aspect of the work will be produced.

5. ANALYSIS

In the research brief (assignment remit) reference is made both to 'the framework for analysis' and to 'the mode of analysis'. These are not identical. (Appendix G)

5.1. Modes of analysis

Methodological texts give lists of modes of analysis. (Denzin and Lincoln, 1994 Part IV) This research will mainly deploy content analysis and thematic analysis, as these are conventionally defined.

In an attempt for greater transparency, three personal definitions of three modes of analysis to be used are now offered. These seem applicable to both the minor and major strands.

- | | |
|------------------------|---|
| Mode 1 Analysis | This will put the research questions in the foreground of attention. The data will be dissected in a search for answers. |
| Mode 2 Analysis | This will emphasis etic issues and imported concepts and will reflect the pre-occupations of the researcher. The data will be interrogated to elucidate issues and to clarify concepts and their inter-relationships. |
| Mode 3 Analysis | This will try to identify emic issues and emerging concepts. The data will be scrutinised to reveal concepts and relationships which are important within the perspectives of those researched and which are not anticipated by the researcher. |

5.2 Structures for analysis

Such fragments of a conceptual/theoretical framework as are given here should be sufficient to guide the overall research design, the separate methods and some of the analysis. It is, however, argued that one purpose of the analysis (served by modes 2 and 3) is to achieve some measure of integration of the fragments and develop a framework which is more complete and satisfying in ways that cannot be foreseen. What are these fragments? What big ideas do they incorporate and what relationships do they suggest?

- (a) There are concepts and relationships which spring from the description of context in Section 1. The following fuzzy generalisations (Bassey, 1999 - p 54) imply a loose contingency rather than a direct causality.

There are many factors influencing change in how students are assessed, only one of which is emerging assessment policy.

There are many features of developing policy, only some of which directly influence how students are assessed.

Changes in assessment practice have an impact on both traditional and new conceptions of professionalism.

Emerging conceptions of professionalism can reinstate assessment as an enhancer of student learning.

Current policy-making undermines traditional conceptions of professionalism.

Emerging conceptions of professionalism can result in more focused policy activity.

(b) There are big concepts which currently seem significant to the proposer. Here a fevered brain succumbs to compulsive alliteratitus. (Perhaps, in this month of McCrone, there is a need for CAs - chartered assessors?)

Constructive alignment features increasingly in both the specialist assessment literature and in policy-related texts. (Biggs, 1995 b and Jackson and Lund, 2000) The notion applies to the relationship between learning outcomes, teaching methods and assessment procedures; it could be extended to cover the relationships amongst national, institutional, faculty and departmental policies. When constructive means effective and when alignment is a policy directive, then constructive alignment can have malign aspects; when constructive means creative and alignment is a purposeful activity of the practitioner, then constructive alignment is likely to be benign.

Constrained autonomy is seen as a salient feature of recent policy-making, but it has always been a prerequisite for justice in assessment. The nature of both the constraints and the autonomy deserve thorough analysis.

Competent assessment is an explicit goal of some recent policy-making. Within existing practice there is undoubtedly some incompetence and, wherever it occurs, it is intolerable. However competence is a necessary but not sufficient condition for capability in assessment; mere competence is a mean and inadequate goal.

Creative agency is required of university teachers both in assessment practice and policy activity.

(c) There are two bigger ideas than these; for these the proposer is pleased to return to Aristotle (as reinterpreted by Dunne, 1993). The concepts are those of *techné* (technical rationality) and *phronesis* (practical knowledge). In a final chapter assessment policy and practice will be analysed with reference to these two fundamental concepts. The analysis may focus on the following.

Constructive alignment in policy-making is an imposition of technical rationality; constructive alignment by practitioners requires practical knowledge.

*Competence in assessment depends on *techné*; capability in assessment requires *phronesis*.*

*Autonomy and creative agency are constrained by *techné*; their reclamation is through *phronesis*.*

Policy-making as techné enfeebles traditional professionalism; a new professionalism requires the would-be professionals to rediscover phronesis.

Current policy exalts techné; future practice deserves phronesis.

CH June 2000

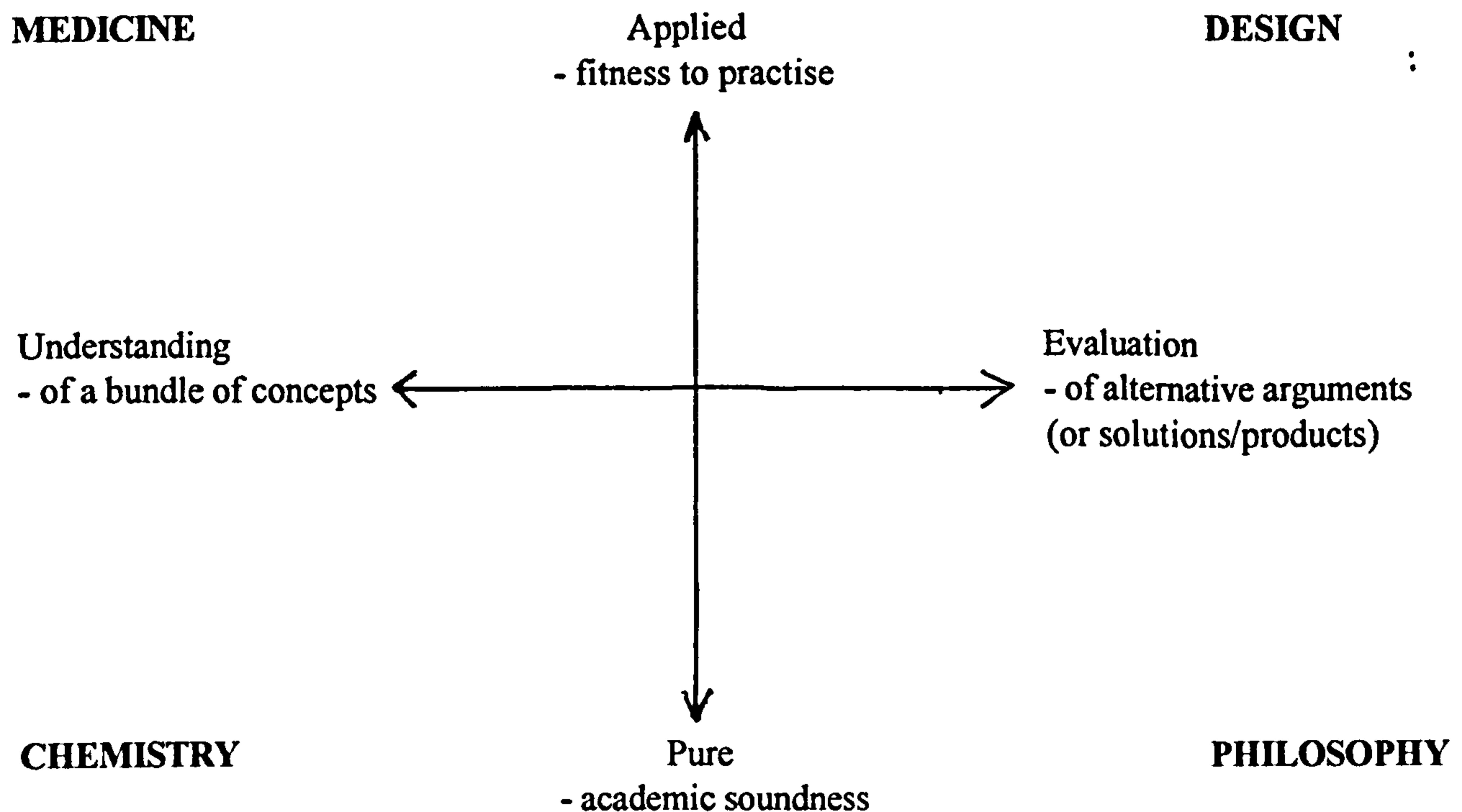
REFERENCES

- Bassey, M. (1999) *Case Study Research in Educational Settings*, Buckingham: Open University Press.
- Biggs, J. (1999 a) *Teaching for Quality Learning at University*, Buckingham: The Society for Research into Higher education and Open University Press.
- Biggs, J. (1999 b) 'What the Student Does: Teaching for Enhanced Learning', *Higher Education Research and Development*, 18(1): 57-75.
- Bottery, M. (1998) *Professionals and Policy Management Strategy in a Competitive World*, London: Cassell.
- Boud, D. (1995) and Brew, A. (1995) *Enhancing Learning through Self-Assessment*, London: Kogan Page.
- Bridges, P., Bourdillon, B., Collymore, D., Cooper, A., Fox, M., Haines, C., Turner, D., Woolf, H. and Yorke, M. (1999) 'Discipline-related Marking Behaviour Using Percentages: a potential cause of inequity in assessment', *Assessment and Evaluation in Higher Education*, 24(3): 285-300.
- Brown, G., Bull, J. and Pendlebury, M. (1997) *Assessing Student Learning in Higher Education*, Routledge: London.
- Brown, S. and Glasner, A (eds) (1999) *Assessment Matters in Higher Education*, Buckingham: The Society for Research into Higher Education and Open University Press.
- Colebatch, H.K. (1998) *Policy*, Buckingham: Open University Press.
- Cowan, J. (1998) *On Becoming an Innovative University Teacher*, Buckingham: The Society for Research into Higher Education and Open University Press.
- Crotty, M. (1998) *The Foundations of Social Research: Meaning and Perspective in the Research Process*, London: Sage.
- Denzin, N.K. and Lincoln, Y.S. (eds) (1994) *Handbook of Qualitative Research*, London: Sage.
- Dunne, J. (1993) *Back to the Rough Ground: Practical Judgment and the Lure of Technique*, Indiana: University of Notre Dame Press.
- Erwin, T.D. and Knight, P. (1995) 'A transatlantic view of assessment and quality in higher education', *Quality in Higher Education*, 1(2): 179-188.

- Freeman, R and Lewis, R. (1998) *Planning and Implementing Assessment*, London: Kogan Page.
- Freidson, E. (1994) *Professionalism Reborn: Theory, Prophecy and Policy*, Cambridge: Eliot Freidson.
- Hammersley, M. (1998) 'The Profession of a Methodological Purist' in Walford, G. (ed) *Doing Research about Education*, London: Falmer.
- Holroyd, C. (2000) 'Are assessors professional? Student assessment and the professionalism of academics' *Active Learning in Higher Education*, 1(1): 62-78.
- HEQC (1997) *Assessment in Higher Education and the Role of 'Graduateness'*, London: Higher Education Quality Council.
- Jackson, N. and Lund, H. (eds) (2000) *Benchmarking for Higher Education*, Buckingham: The Society for Research into Higher Education and Open University Press.
- Miller, A.H., Imrie, B.W. and Cox, K. (1998) *Student Assessment in Higher Education*, London: Kogan Page.
- NCIHE (1997) *Higher Education and the Learning Society: National Committee of Inquiry into Higher Education. (The Dearing Report)* London: NCIHE.
- Nixon, J., Martin, J., McKeown, P. and Ranson, S. (1997) 'Towards a Learning Profession: Changing Codes of Occupational Practice within the New Management of Education', *British Journal of Sociology of Education*, 18(1): 5-28.
- Ozga, J. (2000) *Policy Research in Educational Settings: Contested Terrain*, Buckingham: Open University Press.
- QAA (1999) *Code of Practice for the Assurance of Academic Quality and Standards in Higher Education: External Examining - Draft Guidelines*, London: The Quality Assurance Agency for Higher Education.
- QAA (2000) *Draft Code of Practice for the Assessment of Students*, London: The Quality Assurance Agency for Higher Education.
- Rowntree, D. (1987) *Assessing Students - How Shall We Know Them?* London: Harper and Row.
- Stake, R.E. (1995) *The Art of Case Study Research*, London: Sage.
- Toohy, S. (1999) *Designing Courses for Higher Education*, Buckingham: The Society for Research into Higher Education and Open University Press.

APPENDIX B

Choosing subject area sites



The above hypothesises that assessment (at Final or Honours stage) will be described in terms of two dimensions of candidate performance: (i) a 'pure to applied' dimension and (ii) an 'understanding of key concepts to evaluation of arguments/solutions/products' dimension.

Medicine is predicted to emphasise fitness to practise and understanding of key concepts.

Design is predicted to emphasise fitness to practise and the ability to evaluate opposing solutions/products.

Chemistry is predicted to emphasise academic soundness and understanding of key concepts.

Philosophy is predicted to emphasise academic soundness and the ability to evaluate alternative arguments.

The applied pole reflects a vocational or professional preparation purpose. This is not the same as the introduction of a transferable skills element into academic courses to enhance employability.

PROPOSED RESEARCH ON METHODS OF ASSESSMENT

Aspects of Approval and Anonymity

1. The purposes of the research are all related to description, understanding and illumination; there is no purpose related to evaluation or criticism. The focus is exclusively on the rationales for different assessment methods.
2. In any report, oral or written, arising from the research, the following rules will be observed.

(i) No individual member of staff will be named.

Any view directly quoted will be attributed to a member of staff assigned a number.

(ii) Subject areas will be named.

The research will only take place in departments where the head of department agrees to the naming of the subject area.

(iii) The University of Glasgow will not be named.

Although The University will not be named, it should be noted that it may be identifiable through a determined study of, say, acknowledgements and references.

3. The above rules will allow statements of the following types to be made.

Five of the eight members of staff working in this subject area said that change had occurred because....

'This was done in an attempt to reduce the assessment burden on both staff and students.....' (Staff member 6)

Agreement to seek interviews with seven members of staff of the chemistry department was obtained from the head of that department.

The interviews were conducted with staff from four subject areas in a university in the United Kingdom; it is referred to from now on as The University.

4. All people with whom interviews are sought will be sent a copy of this protocol and a letter which outlines
 - (a) the purposes of the research
 - (b) the main topics to be covered in interview
 - (c) an estimate of the probable length of the interview.
5. After all the interviews in a subject area have been completed, a draft of the full chapter reporting findings from that area will be sent to each person interviewed (and to the head of department if s/he was not an interviewee), providing an opportunity for him or her
 - (a) to correct errors of fact
 - (b) to indicate any disagreement in matters of interpretation.

A summary of the final version of this chapter will be sent to the head of department for distribution as s/he believes appropriate.

APPENDIX D

Some general points from informal conversations (March 2000)

1. I have general approval in principle to proceed.

The conversation with the Vice-Principal was highly satisfactory in one way: 'That's fine; most interesting; of course you have my approval; best wishes.' In another way it was disappointing. I had the distinct impression that his mind was on his next meeting, which was probably of much greater importance to him. I wanted a serious engagement of minds; what occurred was an amiable but perfunctory chat. It does seem as if he trusts me to get on with things and not deal with departments in ways likely to raise problems that would force his involvement.

Members of senior management are under pressure to understand all the important issues with which they may be faced, but clearly they cannot engage in everything and have to prioritise. The role of senior management in relation to assessment is an interesting one and perhaps deserves a research focus of its own. This particular senior manager, to judge from remarks on this and other occasions, has some ambivalence about assessment: on the one hand, he does not see it as of great importance and is happy to leave it as the proper responsibility of departments and the registry; on the other hand he is well aware of increasing national pressures to produce a code of assessment in conformity with the demands of national policy.

The Director of TLS also gave his approval. This was a much more nourishing encounter, but if anything I think he could have been more critical. 'I would be likely to approve of any research proposal you came up with; I trust you not to make a mess.' The re-occurrence of the matter of trust is intriguing. Whereas the Vice-Principal appeared to trust my diplomatic sense, the TLS Director seemed to be speaking of trust in my presumed research expertise. I take these mentions not to be mere flattery, but as a warning that I will have to engender both forms of trust in those who do not already know me. This will need to be done through care in the procedures for securing access to people, in what I write to people and in how I behave in the interviews (particularly in their opening stages). I have to earn trust and then remain worthy of it.

2. Colleagues in TLS expect more of the research than it can deliver.

TLS colleagues all said they found the research outline of great interest and urged me enthusiastically to proceed. Despite my stress on its modest scope, they seemed to think the product would be of interest to the university as a whole, that it would encourage more creative thinking about assessment practice and that it would inform future planning and provision of support on assessment from TLS. Their acceptance of my research plans was based on a practical work-related perception of possible outcomes.

I am not overly modest, but I think their expectations are unreasonable. The likelihood of research

of this nature having direct positive applicability to practice is small.

3. Other people were initially sceptical, if not apprehensive.

In the first stages of our conversations, the departmental people were all polite, but a little cool. It was pointed out that there was very little tradition of this kind of research into higher education in the University of Glasgow; two people said that they quite regularly received research questionnaires from outwith the university and the tendency was always to chuck these in the bin. They could not predict how a research request from inside Glasgow would be received. (I have a hunch that people will find it harder to refuse a personal approach to be interviewed than a questionnaire from an unknown source. This is partly due to how the research methods themselves are perceived.)

Although very explicitly stressed that my purposes were those of illumination and not criticism and that I was interested in the rationales for assessing students in the ways we do and not in instances where the operation of procedures might be said to have gone wrong, there was clearly a worry that I might uncover things which if published would do damage to the department and possibly to the university as a whole. If there seemed to be a lack of trust in the procedures and indeed the value of educational research, there seemed to be a corresponding lack of confidence in their own practice and in their own thinking about it. 'Assessment just gets done; I am not sure that people will have thought deeply about why.'

As the conversations proceeded people became noticeably more engaged and positive and all concluded by saying we had had 'a most interesting talk'. It may be that most academics are likely to become interested in anything that is discussed at some depth. It is of course possible that initial feelings of unease could be dissipated during interview and then revived after the discussion was over. Initial scepticism could be assuaged and then reappear as hostility on the publication of a report.

4. People expressed uncertainty about whether I would be received positively by potential interviewees.

It was interesting that everyone said, without any prompting, that they just could not predict how individuals would react to a request from me to be interviewed. 'Most people are decent and helpful - they will respond positively and generously with their time if approached in the right way. On the other hand I do think you are likely to meet a kind of defensive negativity - or indeed paranoia.'

Everybody said that I should be aware of the current constraints and pressures in academic life, in particular about the increasing demands made on the time of academics. 'Don't ask too much time from them - and don't ask for time at the wrong time.' The majority cautioned me against using term 3: 'People are too busy doing assessment to think about it!'

5. People were undecided about whether or not my own cautiousness was justified.

Most people expressed appreciation for my sensitivity to, and explication of, tricky issues and potential problems. However, while recognising my wish to be open and transparent about aspects

of approval and anonymity, they saw a risk that I would prompt people to think about problems that would otherwise never have occurred to them. 'You really are being oversensitive and cautious. Don't try to think of every conceivable problem in advance. Get it roughly right and then just go for it. Snags will arise on the way, but nothing you can't handle.' (By this time I was beginning to feel oppressed by the burden of trust placed upon me.)

6. Some departments will probably co-operate; others may not.

Three people recommended that I have two or three options within each of my categories of subjects, so that I would not be too disconcerted if any particular department refused to cooperate. This good advice should be accepted.

7. Everyone thought Model B for the choice of sites (two axes and four cells) was more interesting than Model A (positions on a hard-soft spectrum).

I now think this model should be adopted but its assumptions should become part of the research.

8. Suggestions for the choice of people within each site were favourable received.

Everyone stressed that the structures and terminology were different from one department to another, but that I had got the type of people right. Most questioned the need to involve any deans or assistant/associate deans in any way at all, whether as gatekeepers or actual interviewees: 'It's heads of departments who matter here.' I was well-warned that 'deans and heads of departments are very busy people and may well shunt you on to someone else'.

I think I should risk ignoring the faculty level as permission-providers even although I have a hunch that although deans do not see themselves as 'responsible' for what goes on in departments, they do not like to be surprised to discover that something like this is happening within them. Perhaps I shall inform them as a courtesy what heads of departments have agreed to - when they have. I will retain the possibility of interviewing faculty representatives because the policy dimensions of the study require attention to the faculty as a source of policy affecting departmental assessment practice. A decision on this can be deferred.

9. The procedure for gaining initial access was commended.

No-one had any objections or critical comments of any kind on the suggested procedure, except in respect of the dean (see above).

10. There was no consensus view on anonymity.

The draft protocol taken into six of the informal conversations had optional suggestions in the anonymity section. There was lengthy discussion of the issues involved; these discussions were fascinating and could be reported here at great length; they provide interesting materials for teaching

materials for a research methods course. There were conflicting views and no answers emerged: 'You'll just have to work out your own salvation.' (My own preferred version was used in the actual access negotiations and appears as Appendix C.)

It was interesting that three people saw the University of Stirling as some kind of problem here. 'It would be OK if you were just doing it in Glasgow and for Glasgow - but it really does worry me that someone in Stirling may get a good laugh at our expense. I see your involvement with Stirling as a genuine problem. Why did you go there in the first place? And when you did, why did you not then decide to do the research on Stirling? I suggest you seek an embargo on the thesis.'

One person strongly recommended that what I must do is come up with a set of rules which would be appropriate for publication of an academic article; those rules could then automatically be adopted within the University of Glasgow. I should not agree something appropriate for limited dissemination in the first instance and then have to renegotiate, and perhaps rewrite, for a wider audience. This recommendation I will try to follow.

11. The idea of a written protocol was favoured by all.

The production of the protocol was seen as a good thing: 'I find it very reassuring. If people don't already know you, it will help them to trust you'. However, as indicated above, people did not want too much detail. 'It's a bit like Nick Ross on Crimewatch. You scare the living daylights out of people and then tell them it won't happen to them and to sleep well at night.'

Additional note

Perhaps there is a problem with choosing medicine. (This was not discussed with others.) This is the only one of the four where I know in advance that there have been huge changes in assessment procedures. This makes it more interesting; on the other hand the answer to questions about change in assessment methods can be given very simply: 'Because we redesigned the whole course'. But there are still fascinating questions about where policy comes into it - some is from the 1993 report 'Tomorrow's Doctors' and some is not - and why these particular methods of assessing students were chosen. The new medical curriculum is of wide interest well beyond Glasgow and there is a particular attraction in studying the assessment aspect of it. Access to medicine will probably be easy and I think I know the right people. The obvious alternative to medicine is engineering; my hunch is that there hasn't been any change there, but there ought to have been. I have some contacts, but I think access negotiations would be difficult.

INTERVIEW SCHEDULE

Subject Area X

Points in preamble

- (i) Thanks for agreeing to talk with me. Any questions from my letter and protocol?
- (ii) My purpose is..... NB. Rationale for practice – not operational glitches.
- (iii) May I have your permission to record this? For my convenience only...
- (iv) This should take about xx minutes. Have you a deadline?
- (v) Your current post/role in X?
- (vi) Anything from you?

A. Can we check that I have the current assessment procedures correct? Stop me if I get anything wrong.

(Summary from prior study of relevant course documentation.)

Level 1 X. Honours X.

B. Tell me, please, what assessment you yourself are currently involved in.

Level 1 X. Honours X.

C. You have been assessing X students for how long?

Look back over this time. What changes, if any, have there been in how X students are assessed? Level 1 Honours

Confirm changes a, b, c. Which of these do you think is the most significant?

In your view what were the reasons for this change?

Have assessment changes had a significant impact on your work as an academic? In what ways....?

Have they had an impact in general on the work of your colleagues in X?

D. In X do you think you may be required to make other changes in assessment in the near future? What changes?

Would you personally welcome these changes?

What are the reasons that make it likely these changes will be required?

How will these changes impact on the work of academics?

E. That was about changes you may be required to make. Imagine for a moment that the conditions could be made favourable, are there changes to assessment practice that you personally would like to see? What are they?

Why would you like to see these changes?

What are the conditions that would be necessary to make the changes successful?

F. You have given me the reasons for assessing students as we do, and for changes in assessment practice. I'd very much like to check your reaction to some of the reasons that others give me. There looks an awful lot of this. It's tick-the box-stiff – don't think about your responses too deeply!

Sheet 1	Main purposes of assessment in X.		
Sheet 2	Reasons for change already made	Level 1	
Sheet 3	Reasons for changes already made	Honours	
Sheet 4	Reasons for changes predicted as likely in future	Level 1	
Sheet 5	Reasons for changes predicted as likely in future	Honours	

G. Is there anything else to do with assessment in X you would like to tell me about?

I am most grateful; thank you very much.

Strongly
agree

Agree

Disagree

Strongly
disagree

At Honours stage in X, a main
purpose of assessment is to check

'academic soundness'

'fitness to practise'

understanding of key concepts

ability to evaluate critically alternative
arguments or solutions

SHEET 2**REASONS FOR PREVIOUS ASSESSMENT CHANGE****LEVEL 1**

	Very important for us	Of some importance	Of no importance for us
There was a general feeling amongst staff that change was necessary	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
One 'enthusiast' on the staff persuaded us	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Students exerted a pressure for change (e.g. via staff/student committee)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
An external examiner pressed for change	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Increased number of students made assessment change desirable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Increased diversity within student population made change desirable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Change in course structure made assessment change desirable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Faculty policy on assessment required us to change	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
University policy on assessment required us to change	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
National assessment policy made change necessary	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Student learning could be enhanced by change in assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Awareness of trends in assessing X elsewhere exerted an influence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Educational/ assessment theory exerted an influence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The existence of appropriate assessment technology exerted an influence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

SHEET 3**REASONS FOR PREVIOUS ASSESSMENT CHANGE****HONOURS**

	Very important for us	Of some importance	Of no importance for us
There was a general feeling amongst staff that change was necessary	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
One 'enthusiast' on the staff persuaded us	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Students exerted a pressure for change (e.g. via staff/student committee)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
An external examiner pressed for change	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Increased number of students made assessment change desirable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Increased diversity within student population made change desirable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Change in course structure made assessment change desirable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Faculty policy on assessment required us to change	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
University policy on assessment required us to change	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
National assessment policy made change necessary	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Student learning could be enhanced by change in assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Awareness of trends in assessing X elsewhere exerted an influence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Educational/ assessment theory exerted an influence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The existence of appropriate assessment technology exerted an influence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

SHEET 4 REASONS FOR FUTURE ASSESSMENT CHANGE LEVEL 1

	Very important for us	Of some importance	Of no importance for us
There will be a general feeling amongst staff that change is necessary	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
One 'enthusiast' on the staff will persuade us	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Students will exert a pressure for change (e.g. via staff/student committee)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
An external examiner will press for change	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Increased number of students will make assessment change desirable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Increased diversity within student population will make change desirable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Change in course structure will make assessment change desirable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Faculty policy on assessment will require us to change	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
University policy on assessment will require us to change	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
National assessment policy will make change necessary	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Student learning will be enhanced by change in assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Awareness of trends in assessing X elsewhere will exert an influence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Educational/ assessment theory will exert an influence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The existence of appropriate assessment technology will exert an influence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

SHEET 5 REASONS FOR FUTURE ASSESSMENT CHANGE HONOURS

	Very important for us	Of some importance	Of no importance for us
There will be a general feeling amongst staff that change is necessary	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
One 'enthusiast' on the staff will persuade us	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Students will exert a pressure for change (e.g. via staff/student committee)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
An external examiner will press for change	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Increased number of students will make assessment change desirable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Increased diversity within student population will make change desirable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Change in course structure will make assessment change desirable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Faculty policy on assessment will require us to change	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
University policy on assessment will require us to change	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
National assessment policy will make change necessary	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Student learning will be enhanced by change in assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Awareness of trends in assessing X elsewhere will exert an influence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Educational/ assessment theory will exert an influence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The existence of appropriate assessment technology will exert an influence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Thoughts from pilot interviews

* Summary version 2

1. Duncan, Andrea and John all agreed to recording of the interviews. The interviews lasted 35, 45 and 65 minutes respectively.

2. All the interactions were enjoyed by C; the Ss seemed happy enough – there were no awkwardnesses of any kind. Ss had to think quite a bit; the answers were not glib; there seemed to be little production of a familiar rhetoric off the top of the head. However, C had an odd feeling that although all the Ss were interesting, they were capable of being more interesting. Perhaps C was asking questions about the wrong topic (i.e. wrong for the Ss.)

Consider an alternative opener for the interview proper. 'What seems to you really good about the assessment of your subject?' Would this have the advantage of encouraging Ss to talk about things that mattered to them? In addition, would reinforce the message that I had no interest in picking fault with current procedures?

3. (Preamble) OK. What should be said about duration? Perhaps: 'We can do this in half-an-hour – but it depends on how much you want to say.'

4. A much longer schedule is possible. This version could be supplemented by a number of prompts (to encourage talk at greater length) and probes (to get further below the surface). With these Ss, prompts are probably unnecessary and C should be sufficiently experienced as an interviewer to come up with probes appropriate to the separate individuals. The insertion of additional material in the schedule could make the whole exercise appear to be more structured than is intended.

5. The focus has been defined as assessment in a subject area at Level 1 and at Honours Stage. Given the way in which potential interviewees are to be identified, they may be qualified (and wish) to talk about Level 1 or Honours or both. This creates an obvious problem with the length of the interview; it cannot easily be predicted and the coverage C would like could take more time than the Ss were willing or able to give.

Perhaps whether they talk about one or two stages could be left to the Ss to decide? Possibly: 'Is there anything distinctive about the other stage that you would like to tell me about?' This needs more thought.

6. There is no need to check C's understanding of current assessment procedures with more than one person (although it cannot be assumed that the first person interviewed is the most appropriate one to do the checking). There is also the interesting possibility that what Ss think is the assessment procedure does not wholly accord with what is formally recorded as the assessment procedure. Consider the possibility of sending each interviewee in advance a summary of the current procedures asking, 'Do you agree that I have got it right?'

7. There is an intriguing issue about whether or not C should define the period in which he is interested. Is the interest in recent changes? Is there a preconception that the most significant changes in assessment are related to those moves (in certain subject areas) to varying degrees of modularisation that occurred seven (CHECK) years ago?

* The first notes were discussed with the research supervisor. Her questions and her comments have influenced this second version.

The question 'What were the changes 7 years ago and what were the reasons for them' does make the approach appear more positivist than is intended. C is interested in how Ss think about how students are assessed and one of the features of this is the consciousness that Ss have of change, what is the most salient thing in their minds, what their memories are of the reasons that operated at the time. So perhaps the question should be something like, 'In your time assessing students what do you see as the main changes in how student learning is assessed?'

8. The 'change in how students are assessed' was in two cases a quite specific change (introducing an assessed oral presentation and making the objective tests computer marked). In the third case it was a very general and diffuse change over a lengthy period of time (gradually introducing the amount of continuous assessment). Both should be considered relevant; they are both covered by the earlier description of what meaning can be attached to 'how students are assessed' (although it is not so clear they would both have come within the now superseded definition of 'methods' of assessment.)

One person talked of a self-contained change and another of a change which was just one element in a series of related changes. ('We wanted more diversity in assessment so we introduced MCQs; this resulted in too much staff time going on manual marking so we moved over to computer-assisted assessment.')

The concern should be with uncovering what people mean by a change in assessment; this is the phenomenological emphasis. Perhaps some typology of different perceptions of change is starting to emerge: change as a specific isolable event; change as a broad change of emphasis; change as a sequence of incremental steps. We should be sensitive not only to whether or not people like change but also to how they construe its nature. People may have different constructions of the process of change, irrespective of what they think of its desirability or the procedures adopted for bringing it about.

9. When a person has been intimately involved in a recent change in assessment there is a tendency for them to go on at great length about details of procedures, individual assessment items, administrative hassles and inter-personal disputes. It is difficult for C to decide on the spot whether this is all redundant and irrelevant information and if it is to move Ss on without appearing dismissive of something of importance to the individuals.

Qualitative data are being gathered; later the search will be made for what within them is irrelevant (as judged by several criteria). But 'things should not be discarded before you have had a damn good look at them'. 'S is going on at extraordinary length – why is she doing this....' (SB)

10. Ss found it difficult to say much about what might be required in future. This is not something people have given much thought too. In two cases the answer was 'Do you mean something like benchmarking?'

11. They found it much less difficult to talk about changes they would like to see if the conditions were favourable. This may well turn out to be an interesting section of the interview, although describing the necessary conditions seems a little unlikely to go beyond 'more staff and more time'.

Perhaps 10 and 11 should be interchanged? People would then talk about what they would like (the ideal) and what they were likely to have to put up with.

12. The 'tick the boxes' sheets were meant to be handed across to Ss for them to complete themselves. Somehow it seemed much easier for C to tick the boxes for them i.e. to deal with the sheets as an orally administered questionnaire. Fascinatingly this did result in a

lot of additional material; people were by then in a discursive frame of mind and just continued that established type of interview interaction. This is fine but C cannot just administer four of these sheets one after the other – it would just be downright tedious. Probably the ‘reasons for required change in the future’ sheets should be abandoned. (Apart from anything else people tended to say things like ‘Who knows? I suppose it might be.... We shall just have to wait and see.’) Perhaps one question will have to do: ‘Which of these reasons will become more important in the future?’

More intractable, however, is the problem of having two of these sheets, one for Level 1 and one for Honours. A decision is needed as to whether to retain two and say ‘We’ve dealt with Level 1 – let’s whip through it again fast for honours’ or only to deal with one sheet and then ask if answers would have been different for the other stage.

There are some minor modifications required to the content of the sheets, but nothing substantial.

13. Sheet 1 (checking the criteria used in the selection of subject sites) may work or it may not. It would probably be better to ask people to record their position on two spectra linking the opposing conceptions i.e. re-design this sheet removing the boxes and actually providing the two axes described in Appendix B. People will not mark themselves at polar extremes, but this is perfectly reasonable given that most final assessments probably do balance the alternative features.

14. People do say quite a lot that is related to emerging theoretical concerns which are not made explicit in the questions. For example there are comments about communicating with colleagues and the necessity for teamwork (?redefined professionalism), about relationships between teaching methods and assessment procedures (constructive alignment), it being right for departments to be different (autonomy) and about consistency and authenticity.

15. A collection has been started of vivid verbatim quotations which might be of use later.

16. This was a pilot not merely of an instrument but of a person, the interviewer. Copious notes were made during the pilot relating to interview technique. These will be considered and taken into account where possible. They will not be reproduced in detail here. Just six comments – and here use of the first person seems desirable.

(a) I am clear what I would like interviewees to think at the end of the interview. There are a few ways I can develop my technique to make this a little more likely.

‘Well, that wasn’t a waste of time. At first I thought Colin was a bit simple-minded, but then I realised probably he wasn’t. He certainly got me to talk – I wonder if I talked too much? I don’t think I’ve said anything I’ll live to regret. Interesting how you don’t always know what you think about something until you hear yourself talking. It’s a bit unusual for people to listen to me carefully for so long at a time; it’s really rather pleasing. Yes, that was all right.... Interesting he comes from TLS; I’d like to know more about that service.....’

(b) As always, listening to such tapes I said, every now and again, ‘You stupid, stupid man. There was something potentially really interesting there which you should have spotted and probed... Do keep your wits about you.’ I am told everyone except the most arrogant does this.

- (c) I manage to avoid crass examples of leading questions. However, when S says 'I'm not sure what you mean, give me an example...' I give an example and it would then be possible to re-cast the episode as a response to a leading question.
- (d) I am much better than I used to be at keeping my mouth shut and making positive use of silence. (I have trained a range of different types of interviewers and if there is one common fault it is that they all talk too much due to an inability to tolerate silence.)

It is clear from the interviews (a) how important open questions are and (b) how the use of a gentle imperative ('Tell me more about that...') can be more productive than the use of any question.

- (e) I wonder if there was too much laughter in these interviews. No-body says anything which is funny, but there is much signalling of shared humanity. This must be a good thing.
- (f) The interviewees in this pilot exercise were all known to me in advance. Perhaps unconsciously I chose people I thought it would be easy and pleasant to deal with. My interview skills were adequate to make the interviews productive with these people. I know there will be people in the research proper who are not easy to work with, indeed I have already been warned by one head of department that there is at least one person on the list who will be 'downright difficult to handle'.

CONCLUSION Some further thought is required on some aspects. The schedule needs a number of revisions, but can be made to work. The interviewer can 'get by' but there is no room for complacency.

SOME MUSINGS ON ANALYSIS

1. Excerpts from documentation

- (i) In the original description of the assignment remit, section 4 was to cover 'frameworks and methods of analysis'.
- (ii) In the revised document, the assignment was required to be 'A research proposal outlining..... the mode of analysis'.
- (iii) The assessment criterion related to (ii) above was given as '..... the adequacy of the framework for analysis'.

Perhaps the assessment criterion could be usefully expanded to '.... the adequacy of the framework for analysis and the appropriateness of the mode of analysis to that framework'.

2. Useful quotes from the working weekend

The framework for analysis	<p>identifies what you are interested in – the big ideas and important concepts,</p> <p>shows why you are interested in what you say you are interested in; it justifies what you are doing,</p> <p>identifies the field and suggests what is worth looking at within the field,</p> <p>provides a clarification of the conceptual focus,</p> <p>influences how you will go on to analyse the data,</p> <p>is necessary to clarify those ideas which are already there although only half-formulated,</p> <p>makes explicit your 'angle' on the research, but note that different theoretical perspectives make different angles more or less likely.</p>
----------------------------	---

'Your research will not be ethnography in some noble anthropological tradition which implies there are no pre-conceptions. You will deliberately have adopted pre-conceptions, of which you are aware; your pre-conceptions provide a temporary framework. That framework will be changed in the course of the research – but it is necessary to have one to begin with.'

3. Analysis as process and as structure.

Simple dictionary definitions suggest that a mode is a way or a manner in which something is done and that a framework is an essential supporting structure. Conventionally if a framework is a structure then analysis is a process. In this case, how are structure and process related?

Some methodological texts give lists of 'modes' of analysis. A fairly typical one would be: content analysis; thematic analysis and discourse/narrative analysis. Sometimes it is suggested that these modes are characteristic of different research paradigms: content analysis is at home within positivist and quantitative approaches; thematic analysis within interpretivist and qualitative approaches and discourse/narrative analysis within critical research traditions. (Sometimes what is listed as a 'mode' looks to this writer more like a type of data source to which the mode is applied, for example, textual analysis, document analysis, event structure analysis.) There is a temptation now to explore the relationships between the theoretical perspective of the research and the framework for analysis; this temptation will be resisted.

The word 'framework' can suggest a range of metaphors.

Framework as skeleton

If the framework is seen as a skeleton, then the process of analysis is what adds flesh to the bones. This seems illuminating for endo-skeleton but less so for an exo-skeleton. There is the fragile body of tortoise; let's grow a protective carapace around it? Does the physiological metaphor suggest that flesh and bone develop together, not one before the other, or simply that framework is to process as anatomy is to physiology?

Framework as scaffolding

The scaffolding exists before the process of construction starts; it makes possible the building of something new. This could be related to constructionism as the theoretical perspective. Perhaps the framework for research defines the zone of proximal development in which analysis operates?

Framework as the lines of a tennis court.

The activity (or game) of analysis is played within the lines and according to the rules. The existence of the framework helps to rule things in and out; it assists in the identification of the irrelevant and in data reduction. The game allows an element of 'ludic lightness', although post-modernism might not agree to the prior drawing of straight lines.

Framework as a net

There are knots in the net; these are the nodes where individual strings tie together. In such a framework the quality of the knots and the nature of the linking materials are both important. If a theory is a statement of a relationship between concepts then a theoretical framework is a net of concepts related by links, the nature of which can be made explicit. Analysis can then be seen in two ways: as the activity of fishing in which big ideas are retained in the net and the little ones are allowed to escape or as the activity of creating a new and better net from existing fragments.

The metaphor for structure of a net rather appeals and there does exist a process called network analysis. (Bliss et al, 1983) Essentially this involves the development of an elaborate system of categories by way of classifying qualitative data and preserving the necessary complexity and subtlety of the data being processed. A notational technique is employed to generate net-like structures that show the interdependencies of categories as they are developed. (There is an example of the process in Cohen and Manion, 1995)

None of the above indicates anything much about the nature of the knots or the concepts. Perhaps after identifying these they should be shown in diagrams which look like the fragments of net from which a network can be generated. Perhaps an analysis network for research should resemble a mind-map for learning.

4. A simple, interim answer

Think of applying three modes of analysis to material in text or on tape..

- | | |
|-----------------|--|
| Mode 1 Analysis | Look for data which contribute to answers to the research questions. |
| Mode 2 Analysis | Interrogate the material for data which are relevant to the concepts already half-formulated (or to relationships between them). |
| Mode 3 Analysis | Interrogate the material for data from which some new big idea (concept or relationship) may be induced. |

This should not be taken to prescribe that the three modes of analysis are applied in sequence and then one stops; progression through the modes should be cyclical and iterative.

References

- Bliss, J., Monk, M. and Ogborn, J. (1983) *Qualitative Data Analysis for Educational Research*, Beckenham: Croom Helm.
- Cohen, L. and Manion, L. (1995) *Research Methods in Education*, London: Routledge.

APPENDIX III

THEORY GUIDELINES

In the past I have tried to give students guidance on various theory aspects of their research. Such guidance was usually issued with a health warning: 'Unquestioning absorption of this advice may damage your intellectual health'. Students challenged my advice too rarely. I have wondered on occasion whether some deeply-buried anxiety about my ability to answer questions might have led me to issue guidelines with such authoritative confidence that few dared to ask them.

When I became a student again, I decided I would ask one or two of those questions I wished people had asked me. The tutors looked mildly disconcerted and did not provide much in the way of direct answers. This did not, of course, mean that they were unable to answer my questions, only that they sensibly returned the responsibility for answering them to me.

There follow six theory guidelines. Each is introduced by some of the questions that students should be encouraged to answer for themselves if they are to follow the guidelines with understanding. My answers to these questions do not aspire to provide some comprehensive overview of the grand narratives of educational research; they are my answers briefly expressed, for the purposes of this research.

(a) *Research for a higher degree must go beyond the descriptive.*

(Why? What is to count as beyond description in this study?)

Some forms of research do not, and need not, go beyond description; market research is an example. One could, of course, argue that because such forms do not go beyond description, they are not 'real' research at all. Around some educational research activity there has been lively debate as to whether it deserves the title of research; for example, the work of the Assessment of Performance Unit in England and the Assessment of Achievement Programme in Scotland were regularly criticised for being overwhelmingly descriptive and largely a-theoretical. Be that as it may, it is wholly reasonable to demand that to be rewarded with a higher degree, research must go beyond description. This is because of the expectation and requirement that such research will add something to our understanding of what it is that is being researched into. Description makes it clearer what it is that is there to be understood, but does not of itself contribute to our understanding of it. It is insufficient simply to provide snapshots of things, to describe them as they are; this may add something to the sum of human knowledge but contributes nothing to human understanding, let alone wisdom.

The most obvious thing beyond description is explanation (as the EdD documentation has it 'to explain why things are as they are'). Evaluation (Are things good as they are?) may be a legitimate research purpose; usually prescription (What

is to be done to improve things) is not. If research aims to bring about greater understanding, then the extent of achievement of that aim has to be judged in terms of the credibility of the explanatory claims it provides. Two cautionary notes are in order here. Firstly, to argue that explanation is necessary in academic research is not to dismiss description as unnecessary, only insufficient. There is little point in attempting to explain 'why things are as they are' without some defensible description of what they are. Secondly, all of the foregoing makes too easy a distinction between description and explanation (and also between knowledge and understanding). What can appear to be a simple description may well have a covert element of theoretical explanation. ('Those birds are flying from north to south' looks like a simple descriptive statement. What then is 'Those birds are migrating'?) There exists a grey area inhabited by what Hammersley (1996) has called theoretical descriptions.

(b) *The conceptual basis of the research must be clear.*

(What is to count as a concept in this research? What does the conceptual basis do? Is 'basis' the most helpful term?)

The concept of a concept is relatively straightforward. A concept is a general notion, an abstract idea, a mental picture (or schema) of some group or class of objects or occurrences. A concept can reasonably be said to be possessed when the putative possessor can reliably identify both members and non-members of the generalisation. The conceptual basis of any research is thus that collection of 'big ideas' which provide a foundation or support for it; without such a basis the research would not 'stand up'. This structural metaphor reveals something helpful, but not enough. Concepts form not only a basis, but an organiser and provider of significance. Without some conceptual basis, the findings of research can not be more than an aggregation of specifics; such an aggregate cannot achieve the significance of an abstraction; it cannot even be a useful (concrete) building material, but remains a mere agglomeration of ingredients. The generalisation and abstraction within a concept are what can give some individual bit of research an import beyond its own specifics; if no concepts are involved, any piece of research is isolated from all others. A lack of cumulative significance is then not only a matter open to practical (and political) criticism but one of epistemological inevitability.

To say that research has to have some conceptual basis to give it coherence and some general significance, is not at all the same as saying that research must start with a fully-worked out organiser in which all the concepts and linkages between them are already clear. It is a legitimate purpose of research, as well as a happy outcome, to bring into some relationship concepts that were not previously seen as related. Starting research with a developed conceptual framework may make the research less messy, but it also restricts scope for creativity. The famous dictum from Koestler (1964) on this subject needs only to be slightly misquoted: 'All acts of creative thought involve the bisociation of previously unrelated matrices or concepts'.

This present research does not start off with a tidy, well-developed conceptual framework. On the other hand, it would be wholly misleading to suggest that this research began only with a set of wholly unrelated concepts and no framework whatsoever. The choice of what were to be the central concepts was made on the assumption that some kind of relationship was likely to exist between the concepts and, more than this, there were ‘hunches’ as to what the relationships might be. The research was not conducted to create a conceptual framework where nothing previously existed; rather it took a very hazy framework and attempted to make it a little less hazy and better developed.

The idea of concepts forming a **basis** for research is useful, but it is necessary to ask when the implied metaphor can become misleading. ‘Basis’ can suggest that concepts are fundamental and that the research is built upon and ‘above’ the conceptual foundations. This is meaningful in some temporal sense i.e. there are initial conceptual notions and assumptions which are imported into the subsequent processes of the research. However, the notion of conceptual basis becomes unhelpful if it starts to convey the idea that what is more basic is more ‘real’. What is it that forms the material of the research? Educational and social research is concerned with events and developments, and in particular with people, what they know, what they think and feel, how they act and react; these are the ‘givens’ for the researcher. The data of research are then analysed and interpreted by the researcher into what is essentially a construction formed of general, big ideas (the abstractions we have called concepts) and statements implying linkages between these concepts. The ‘reality’ of people knowing, thinking and behaving is then the basis of the research and the metaphor of a conceptual **superstructure** seems rather more appropriate than that of a conceptual basis.

What am I actually doing here? I am playing with explanations which take the form of analogies or metaphors. (See page 13) I am suggesting that more than one metaphor is necessary to understand the role of concepts within research in general and this research in particular. We need the metaphors of both supporting basis and interpretive superstructure. It may also be revealing to think of a conceptual shrubbery – a place in which concerns and people get lost.

(c) Theory is taken into research and theory should be developed through the research; both have to be made explicit.

(What kinds of theory are there? What does a theoretical framework in this context look like? How are the theoretical and conceptual frameworks related?)

The word theory can have many meanings attached to it. In education it is currently used to describe (a) a speculative (especially fanciful) vision, (b) that sphere of abstract thought to be distinguished from practice, (c) a statement of values and priorities explicating how things ought to be and (d) a proposition illustrating key principles. There is no space to explore all these here. It is proposed, within this research, to adopt only one ‘strong’ meaning of theory: a theory is taken to be a statement which expresses an ordered relationship between concepts and which has some explanatory value.

What can a framework look like? Sometimes the word framework is used somewhat loosely as a synonym for angle of vision, or perspective or 'work within the school of'; this is obvious when the research approach is modelled on how the researcher believes some particular 'great thinker' would wish him or her to proceed. For example, 'I am using a Foucauldian framework' can be translated as 'In my work I use a set of concepts as these were developed by Foucault'.

The term framework is used in other ways. Firstly, it may mean no more than a frame. Here, it defines the boundaries of the research, defining what is to be included within the picture and what will remain outside. Secondly, a framework may be seen as an essential supporting structure, rather like a skeleton; then the process of research construction involves 'fleshing-out' i.e. adding meat to the bones. Thirdly, it may be a form of scaffolding (an exo-skeleton?) within which the research product grows in an organic way. (Such scaffolding may also provide a zone of proximal development within which the researcher can develop.) The framework, fourthly, can be rather more like a network. This can be either a two-dimensional net of knots and threads or a three-dimensional lattice complete with occlusions and inclusions of impurities and fracture planes.

In this research the theoretical framework is construed simply as a network in which concepts form the knots and the relationships between the concepts are the threads. The net is incomplete; some knots are missing; some of the threads are invisible; some are clearly weak and others are not as strong as they first appear; the nature of most is obscure. However, despite its deficiencies, the net is useful for a limited period. The purpose of the research is to improve the net a little by modifying the knots and by strengthening some threads, cutting out some and adding others.

I developed this metaphor of theoretical framework prompted by dim memories of the writings of Hempel (1966 –see below); I was then intrigued to discover that something called network theory now exists. Network theory presents a way of representing the content and the structuring of knowledge within a defined domain; systematic network analysis involves the development of an elaborate system of categories (or concepts) to deal with qualitative data and to preserve the essential complexity of the area under investigation. (See, for example, Bliss et al 1983) Network theory as applied in research bears a strong resemblance to cognitive maps as applied in learning; network analysis is similar in its operation to cognitive mapping. It is interesting that this parallel suggests that the adequacy of a theoretical framework should be judged not only by its validity, but also by its utility and 'learnability'. A theoretical framework is, in part, valuable if it is accessible to others and is not so esoteric that only its author derives any pleasure from embracing it.

In educational research, there is another important question about theory. **When** is theory? Or, more helpfully, when is theory known in advance of any empirical activity and then imported into it and when is it unknown (or unacknowledged) at the outset and then emerges from the activity of the research? Only in the former case could one make any claim that the research was about theory/hypothesis testing; only in the latter case could one claim to be generating

grounded theory. Rather than re-capping the well-known arguments in this area (starting from Glaser and Strauss, 1967), the stance of this research will be crisply asserted.

- (i) It is impossible to embark on research without some theory; this would require the foolishness of believing in the existence of naïve observation.
- (ii) It is foolish to embark on research which does not aim to develop any theory, however inchoate, which already exists.

There is a pitfall in defining theory in terms of concepts as we have just done; it is tempting to see the concepts as preceding the theory and then the research process forming and modifying linkages that develop the net-framework of theory. It is also the case that generating a better understanding of the nature of the links contributes to a re-defining of the concepts. Hempel put this well many years ago; although he was discussing the philosophy of natural science, his argument holds, with minor changes, for most types of research. Note that in this quotation the word 'superstructure' appears with roughly the same meaning attached to it earlier.

[Research]... requires the establishment of diverse connections between different aspects of the world which are characterised by concepts. The concepts are the knots in a network of inter-relationships in which theoretical principles form the nomic threads. The more threads converging upon, or issuing from, a conceptual knot, the stronger will be its systematic import. The process [of research] might be compared to building a bridge across a river by putting it on pontoons or temporary supports sunk into the river bottom and then using the bridge as a platform for improving and perhaps even shifting the foundations and then again adjusting and expanding the superstructure to develop increasingly well-grounded explanatory systems. Within research, concept formation and theory formation must go hand in hand.

(Hempel, 1966)

How then are the theoretical framework and the conceptual basis (or superstructure) related in this research? The question has almost been answered. The two must be closely related: the conceptual basis/superstructure focuses on the separate knots; the theoretical framework focuses on both knots and threads. Concept formation and theory formation go hand in hand; one is neither epistemologically nor chronologically prior to the other.

This research started with a few broad, fuzzy, multi-faceted concepts and an awareness that there must be some threads linking them; in other words it began not with a theoretical framework, but with some promising theoretical fragments. The net of theory was dimly perceived and seemed both incomplete and in a poor state of repair. The purpose of the research would be to make a modest, but creative, contribution to characterising the threads between the concepts, and through this to a rather better understanding of the component concepts themselves.

Perhaps all this is much more complicated than it needs to be. Suppose this view of conceptual and theoretical frameworks were to be shown diagrammatically, what would it look like? There would be three primary concepts and three relationships, each between a pair of concepts. If the relationships were

asymmetrical, then six statements would be needed to characterise the three relationships.

If these six statements are to assist our understanding they should have some **explanatory** import. Does this mean that they must have the character of fully-formed hypotheses? This would seem unfortunate if it implied that the research was then just being undertaken to test the initial hypotheses; this would suggest a commitment to a thoroughly positivist approach that on other grounds would be uncongenial. The way out of this dilemma depends on what it can mean to **explain** anything. If explanation is only allowed to mean the provision of some simple, automatic, determining causal statement then we have reached an impasse; if explanation is allowed to mean much more than this, then there is a way forward. An exploration of what causality can mean within interpretivist paradigms of research is provided later in this appendix (See page 13)

What is done in this research?

- (a) On the grounds of transparency, accessibility and learnability a simple concept diagram is offered. (See Chapter 2)
- (b) The links between concepts are not given the status of hypotheses. Rather they are seen as simplistic, 'common-sense' assumptions or hunches; they are the kind of tentative possible relationships that anyone might have before embarking on such research.
- (c) It is not assumed that any relationship must be a direct and inevitable cause/effect link; the linkages may have a range of different characters. (See Page 13 of this Appendix.)

What then would be the theory 'taken into the research'?

- (i) The three primary concepts (and two permeating secondary concepts).
- (ii) The 'common-sense' assumptions of linkages.

What then would be theory 'developed through the research'?

- (i) Re-definition of the entry concepts as necessary.
- (ii) Clarification, criticism and modification of the common-sense assumptions about conceptual links.

(d) The theoretical framework for the substance of the research is not to be confused with the theoretical perspective of the research approach; both have to be made clear.

(What theoretical perspective underpins the research approach? What epistemology underlies the perspective? How is the theoretical perspective of the approach related to the conceptual basis and theoretical framework of the content?)

Theoretical perspective and epistemology

There are two obvious questions that have to be addressed in developing a research design. Firstly, what methodology and methods will be employed? Secondly, how is the choice of methodology and methods to be justified? The answer to the second question has to be related to the purposes of the research; the process of the research has to be capable of fulfilling its purposes. This is necessary, but not sufficient. Justification of the choice of methodology and methods reaches into the assumptions about reality that are brought to the research; to ask about these assumptions is to enquire into the theoretical perspective of the research. It also reaches into the understanding we have of what knowledge is, what it entails and what status can be attributed to it. What kind of knowledge will we claim to have been produced by the research? Why should any reader of the research report take the research outcomes seriously? These are epistemological questions.

Crotty (1998) argues that there are really four questions to be answered rather than two.

- What **methods** do we propose to use?
- What **methodology** governs the choice and use of these methods?
- What **theoretical perspective** lies behind the methodology in question?
- What **epistemology** informs the theoretical perspective?

The four elements in bold need a little elaboration.

Methods: the techniques or procedures to gather and analyse data related to the research questions or hypotheses.

Methodology: the general approach, strategy, plan of action (or perhaps paradigm?) lying behind the particular methods and linking methods to desired outcomes.

Theoretical perspective: the philosophical stance informing the methodology and providing a context for the research process and for grounding its logic and criteria.

Epistemology: the theory of knowledge embedded in the theoretical perspective.

For this research how would I answer Crotty's questions? Let's take them in reverse order.

I do not believe that meaning and meaningful reality exist as such apart from the operation of any consciousness, thus my epistemology cannot be objectivist. On the other hand I do not believe that objects as such make no contribution to the generation of meaning i.e. that meaning is only imposed on the object by the subject; thus my epistemology cannot be subjectivist. (I am not a post-modernist.) I do believe there is no objective truth out there waiting for us to discover it; truth and meaning come into existence out of our engagement with the realities in our world i.e. they come from an interaction between subject and object. My fundamental epistemology is thus constructionist.

Someone may now ask what this means for how I view reality; this is a question about ontology (what is, i.e. the structure of reality as such) rather than epistemology (what it means to know). The short answer is that I accept that there is a world and things in that world which exist independently of my consciousness of them, but this does not imply that meanings exist independently of consciousness. The world exists without mind, but meaning does not exist without mind. Crotty concludes that 'realism in ontology and constructionism in epistemology turn out to be quite compatible'. Crude, naïve realism is not consistent with constructionism, but a more 'subtle' realism is.

My theoretical perspective is not positivist, even within those intellectual territories where positivism was once most at home. Even physical scientists now tend to be post-positivist. (There is a weak, purely chronological sense in which most theoretical perspectives are post-positivist; it is necessary to distinguish those which are based on a re-definition or rejection of positivism and those which merely come after it.) In respect of this study of a social life-world, my perspective is interpretivist in that I will be seeking culturally derived and historically situated interpretations. It would be easier for me just to leave it at that, but perhaps someone will probe whether my brand of interpretivism is symbolic interactionism, phenomenology or hermeneutics. The simple answer here is 'Symbolic interactionism – but with some reservation'.

Blumer (1969) explicated three essential interactionist assumptions. Firstly, human beings act towards things on the basis of the meanings that these things have for them. Secondly, the meaning of such things arises out of the social interaction that one has with one's fellows. Thirdly, those meanings are handled in, and modified through, an interpretative process used by the person in dealing with the things encountered. I find myself in complete agreement with these. What is the reservation?

Symbolic interactionism puts great store on the importance of culture. The *mélange* of cultures and sub-cultures in which an individual operates provides him or her directly with a whole world of meanings; it is almost as if the cultural heritage pre-empts the task of meaning-making. There is, I think, a tendency for symbolic interactionism to slide from cultural provision of meaning to cultural determination of meaning. I feel a need to stress the extent to which the individual remains free to **construct** meaning within a cultural context and this begins to sound like phenomenology with its insistence that because culture can constrain and oppress as well as liberate, research should be an attempt to recover a fresh perception of existence, one unprejudiced by acculturation. Again as Crotty puts it, phenomenology is about saying 'no' to the meaning system apparently bequeathed to us and attempting 'to see the things themselves'. Thus, if I understand it aright, I must be a symbolic interactionist who retains a belief in the potency, agency and improvisatory scope of **individuals** as actors within a culturally-provided social drama.

If this is the congenial theoretical perspective, what methodology will be adopted? It will be ethnographic, somewhat liberally construed. Ethnography sits conventionally and comfortably within the theoretical perspective of symbolic interactionism, but what is it?

Ethnography is a form of research in which the social settings to be studied, however familiar to the researcher, must be treated as anthropologically strange; and the task is to document the culture – the perspectives and practices – of the people in those settings. The aim is ‘to get inside’ the way each group of people sees the world.

(Hammersley, 1985)

Why will there be a somewhat liberal construction of ethnography? I can make this clear by dwelling a little on grounded theory, sometimes closely linked with ethnography and again dwelling happily within symbolic interactionism. Grounded theory (see Page 5 of this Appendix) is appealing as an account of how theory may emerge or be generated from research data; it is less satisfactory in that it may encourage neglect of imported or extant theory. (Strauss and Corbin (1994 – page 277) noted that Glaser and Strauss had earlier greatly underplayed the role of extant theories and the ‘unquestionable fact and advantage’ that trained researchers are theoretically sensitised.) The idea of any researcher embarking on any study without any theory, however obscure, embryonic and undeveloped, is to me just untenable, even although it may well be helpful for a time to act as if no prior theory existed. If everything is emergent and nothing is imported then there can never be any corroboration or cumulation. This research will attempt to make explicit the theoretical notions taken into the research; this may debar it from ethnography in its purest and most restricted sense.

What research methods will be used? As I said in the first paragraph of this section, these will be methods judged capable of fulfilling the purposes of the research i.e. methods which can provide data which begin to answer the research questions. Theoretical perspective and methodology may govern how the data acquired will be analysed and used, but they do not prescribe methods to gather them. Having said this, the research methods that will be employed are ones that are commonly used within ethnographic approaches: semi-structured interviewing, participant observation and thematic analysis of documents.

You may have expected me to use the terms qualitative and quantitative long before now. The conventional division in most research methods courses between on the one hand objectivist research carried out by quantitative methods and on the other constructionist research carried out by qualitative methods is, to my mind unhelpful (in that it has also resulted in far too many futile time-wasting arguments) and indefensible. There is no theoretical prohibition on one research study being both qualitative and quantitative in different aspects or phases. To avoid tedious parading of the arguments in support of this conclusion, the reader is referred again to Crotty (1998 – pp 14-17) and to Hammersley (1996 - pp 13-23). Many methodologies known today as forms of ‘qualitative’ research have in the past been carried out in an utterly positivist manner; this is certainly true of the early history of ethnography. I am not going to avoid the use of these familiar terms altogether; I just believe their unexamined use as simple descriptors of research approach and method is unhelpful.

Constructionists may make use of quantitative methods, but their constructionism makes a difference in how the data are handled and what kinds of

knowledge claim can legitimately be made. This present research will not use any allegedly 'quantitative' methods, although it will make use of some numbers and of those crypto-quantitative formulations characteristic of qualitative researchers 'most', 'some' and 'few. However, there is one feature of the main research method that does need some elaboration.

It is routinely said that one defining characteristic of qualitative methods is that data are unstructured at the point of collection. This research study will make heavy use of semi-structured interviews. (And note in passing that semi-structured interviews, so common in qualitative methods, are by definition **not** unstructured.) Almost all of each interview will be devoted to trying to get extended accounts of the interviewee's perspective via a conversational process characterised by open questions ('Why, in your view, did that happen?') and the gentlest of imperatives ('Tell me more about that, if you will...'). However, within each interview in the major strand there will be two short episodes which are relatively structured. In the first of these, participants will be presented with a diagram. I will say (a) 'I wonder if this diagram applies in your subject area? Criticise it for me...' and (b) 'Locate your subject somewhere on this diagram.' In the second short episode, interviewees will be presented with a list of fifteen factors said by others to be influential and asked 'In your view, were these influential factors in **your** area?' These two short episodes look quite highly structured and both allow some modest quantitative analysis and reporting in quantitative terms. This somewhat unconventional procedure is justifiable within the overall research approach; there is, however, some need for comment on it in the reflexive section of the thesis.

At a recent EdD seminar aimed at 'explaining what lies beneath research methodology', one speaker had as his title 'I was born under a wandering epistemology'. I identified with him very strongly. The tutor's comment was, 'Be confident in your uncertainties and give us in your thesis the narrative of your methodological journey.' Rather than narrate within the body of the thesis the lengthy story of my own epistemological wanderings, I have tried to describe in this appendix where, for this research, I am methodologically most at home.

It will be obvious from the foregoing that Hammersley has had an influence on where I have chosen to be at home. Just a little of the personal story may be of interest. I worked closely with Martyn Hammersley for thirteen years (1988 – 2000) on the development, teaching and examining of two Open University courses on educational research. (E812 and E835) When I was puzzling over how best to describe my epistemological and methodological home for this research, I wrote to him in the following terms.

You and I have discussed educational research many times and at length. My views are not identical with yours, but you have influenced me a lot. I know you pretty well, but if you were asked to 'label' yourself as a researcher what would you say? What is the theoretical perspective that underpins your usual research approach?

This is the relevant part of his reply.

I think you should bear in mind that most researchers use these labels loosely and they usually qualify them: 'I'm not a post-modernist, but...' and 'While I don't agree with everything Foucault has said....' In fact, I am not at all sure what I'd call myself. I suppose post-positivist is the simplest label. After all no-one would really want to admit to being a positivist or a pre-positivist, so it covers a multitude of sins. In fact it is quite helpfully ambiguous. Some people treat it as referring to people who are not positivist but who have not moved very far (i.e. not far enough) from positivism. Others use it to refer to all positions that 'came after' positivism, including post-modernism. The other label I rather like for myself is 'subtle realist'.
(Hammersley, 2002)

Theoretical perspective and theoretical framework.

At the EdD seminar just referred to, another speaker first described her theoretical perspective as being contained in two quotations from Vygotsky (relating gesture in language teaching to semiotic mediation). She later described her theoretical perspective as having undergone a shift from positivism to post-structuralism. I asked a question. 'Theoretical perspective seems to be about both the substance of your research focus and about the nature of your research methodology. Might it be helpful to distinguish between a theoretical framework that was about the content of your research and a theoretical perspective which was about the process and then seek to understand how the two might be related?' The chairman's response was: 'That is an exceptionally good question – one to which we must return'. As is in the nature of these things, we did not return. Let's try to answer it now.

A few pages back, I fumbled towards a description of theoretical framework to serve the purposes of this research. The theoretical framework was pictured as a network of knots (the big ideas, the organising concepts) and threads (the relationships between the concepts). The concepts were things like, policy, assessment practice and professionalism. In the previous section I summarised Crotty's position of the theoretical perspectives underpinning methodologies and methods. The perspective was pictured as a kind of nest, nested inside epistemology and ontology and having nested within it approaches and methods. The concepts were things like interpretivism, symbolic interactionism and ethnography. So, are we just talking about a somewhat artificial difference between two metaphors, the net and the nest, or is there some more 'real' distinction here? (Or, even, does it matter?)

Think of research focus A; important concepts are, let's say, semiotic mediation and gesticulation. Think of research focus B; important concepts are policy-making and traditional professionalism. This is not meant to imply that different foci must have different and non-overlapping concepts. One simple example: the concept of 'intelligent practice' could well be common to research in both gestured language-learning and higher-education-assessment. There is an infinite number of research foci and concepts. There is not (as yet) a limitless number of research approaches, only a plethora. This is saying little more than that there is a limited number of research methodologies used to tackle a limitless

number of substantive research areas; there is a limited number of theoretical perspectives and a limitless number of theoretical frameworks. It would seem to follow that there is some value in distinguishing the framework and the perspective, at least as we have defined them here.

The theoretical perspective does not determine the areas to which it can be applied; any particular area can be studied by a whole range of approaches. (However, when a focus is more closely defined by research questions, these questions will carry implications for the methods by which they are most likely to be answered.) The perspective of the approach does not determine the concepts of the substance, but it does influence the knowledge claims that can be made about the linkages between concepts.

Some time ago, I asserted that statements of a relationship between concepts were only of much interest if they had some explanatory potency (i.e. contributed to our understanding of why things are as they are). This assertion now needs closer examination. At this point I wish, not for the first time, that I was much better educated in philosophy. What follows attempts to be simple and accessible and thus is likely to be seen by philosophically capable critics as simplistic and unskilled. I am nevertheless encouraged to press on by a remark of Steven Weinberg's: 'The fact that we [researchers] do not know how to state in a way that philosophers would approve what it is that we are doing in searching for explanations, does not mean that we are not doing something worthwhile. We could use some help from professional philosophers in understanding what it is that we are doing, but with or without their help we shall keep on doing it.' (Weinberg, 1993)

What does it mean to **explain** anything? Let's start from the statement that an explanation is any satisfying answer to a 'why' question. First of all, rule out of consideration 'why' questions to be answered in terms of purposes served i.e. teleological questions. Our concern is with 'why' in the sense of why/how did it come about; explanation is seen here as being about causality rather than purpose. However, even when dealing with causality, the word 'why' is notoriously slippery. For example, the philosopher Nagel (1961) gives ten examples of questions in which 'why' is used in ten different explanatory ways from 'Why does ice float in water?' to 'Why did Cassius plot the death of Caesar?' Weinberg himself takes an apparently simple question about the physical world ('Why is chalk white?') and in a brilliant analysis provides answers at ever deeper levels until he reaches the question 'Why does nature obey the principles of relativity and quantum mechanics?' Weinberg's dream of a final theory, to answer all 'why' questions, would be rejected out of hand by many as crude reductionism, daring as he does to imagine that the complexities of social interaction might one day be explained in terms of quarks, leptons and local gauge symmetries. A narrative far too grand would be the post-modern verdict. What may be less easy to dismiss is Weinberg's initially astonishing argument that explanation is a mode of behaviour that gives people pleasure, like love or art. 'The best way to understand the nature of explanation is to experience the peculiar zing that you get when someone (preferably yourself) has succeeded in actually explaining something.' This is my justification for including the word 'satisfying' in the second sentence of this paragraph. An important feature of explanation is that it pleases or satisfies the person asking the 'why' question. This does mean that satisfying is not the same as

satisfactory. Someone could find some explanatory claim personally satisfying, but to be satisfactory it would have to pass scrutiny from those qualified to criticise it. This then becomes a matter of the credibility of the claim, the credibility of the evidence adduced to support it and the relative credibility of the alternative explanations available. Satisfaction is personal; satisfactoriness is a judgement after rigorous peer-criticism.

Explanations then are of different types and exist at different levels. In any Weinbergian type of analysis, the idea of levels is clearly about depth of explanation i.e. about digging deeply to ever more fundamental levels. It is, however, equally plausible to construe level in a rather more horizontal way as width, scope or scale. This is an odd way to put it; an example may be helpful. Why is black, working-class Melanie shouting in the classroom? The explanation may be given in terms of the interaction of this girl at this time with this teacher – a micro-level explanation. Or in terms of her membership of a school-resistant sub-culture – a meso-level explanation. Or in terms of a neo-Marxist theory of class and race struggle – the macro explanation. We begin here more clearly to discern how the theoretical perspective of the researcher makes one type of explanation more likely than another. The types of explanation offered by the phenomenologist are different from those of the critical-feminist.

When offering explanations, researchers with different theoretical perspectives imply or claim different things about causality. For example, and restricting ourselves to the kind of theoretical perspective met in this research, it is often asked if an interpretivist study can establish any causal relationships at all. The possibility is often attacked from both ‘the right’ (‘Only controlled quantitative experiments can do that’) and ‘the left’ (‘Causality is inapplicable in human behaviour; people are not billiard balls’). I take up the position that interpretive, qualitative studies are well-suited to finding causal relationships. In the words of Huberman and Miles (1994), ‘[such] studies can look directly and longitudinally at the local processes underlying a temporal series of events and states, showing how these lead to particular outcomes and ruling out some rival explanations ... we can understand not just that a particular thing happened, but how and why it happened’. The credibility of this stance depends on how one sees the nature of causality. The position is probably incompatible with a concept of causality restricted by the classical criteria of temporal precedence (A before B), constant conjunction (when A, always B) and contiguity of influence (a plausible mechanism links A and B). It is necessary to add beyond these strength of association (much more B with A than with other possible causes), biological gradient (if more A, then more B), coherence (the A-B relationship fits with what else we know about A and B) and analogy or metaphor (A and B resemble the well-established pattern noted in C and D). In addition within this kind of research there is always causal multiplexity; causes are always multiple and conjunctural, combining and affecting each other as well as their supposed effects.

It is interesting to note that some researchers in interpretivist (and post-modern) paradigms get round their distaste for causes by dealing only with contingencies. When pressed to clarify the distinction, they say that contingencies emphasise that particular events and developments do not happen of necessity, for any single reason, but are just one possible outcome of a whole series of complex

relations between other factors and events. When it is admitted that causes can, as indicated above, be 'multiplex and conjunctural', any useful distinction between causes and contingencies would seem to disappear.

What then does it mean to explain anything? It is conceivable that in many situations none of the above is of much practical significance. In these, it may be enough simply to assert that to explain means to help people understand something they find perplexing. Any explanation offered should be both satisfying and satisfactory. The latter depends on the credibility of the evidence made available to support the explanatory claim.

To conclude this section as a whole, it is for some purposes useful to distinguish the substantive theoretical framework from the theoretical perspective of the methodology. However, the theoretical perspective influences the nature of the relationships that can plausibly be claimed to exist within the framework.

INTERPOLATION: A NOTE ON 'CONCEPTUAL AGENCY'

In discussing the types of causal relationship that can be credibly claimed to exist within a theoretical framework, I am aware of something troubling me. I have been unable to find this matter directly addressed in the methodological literature, presumably because I have not formulated the question appropriately or searched for help in the right places. What follows is a wholly personal treatment.

A theoretical framework leads to the making of statements of the form 'concept A is causally related to concept B'. But a concept is an abstraction residing, or not, in the head of a person. How can an abstraction cause anything at all, let alone another abstraction? How can an idea be a causal agent? Let us start by exploring a simple example taken from a field other than higher education.

Researchers studied the teaching of science in primary schools. The research resulted in a claim and conclusion that can be summarised here as 'lack of confidence in understanding scientific ideas resulted in expository teaching'. (Harlen and Holroyd, 1995) This summary statement expresses a relationship between two abstract ideas; it is a theoretical construction arrived at by the researchers. What happened to allow them to write such a thing? Looked at in a forward direction in time, a primary teacher did not feel confident about his/her understanding of some key some scientific ideas and then went on to teach science in one way in preference to others. Inferring causality was of necessity a retrospective matter. The researchers studied the scientific understanding possessed by the teachers and the ways in which they acted when teaching and then from a collection of such cases made a general claim stating a relationship between teachers' confidence in their understanding of scientific ideas and their preferred teaching styles. The research data were about what teachers knew and about how they acted; events and activities were the result of the knowledge, emotions, values and predispositions of people. People were the agents of events. The conceptual/theoretical superstructure was an interpretive construction by researchers; elements of this superstructure did not cause other elements. In this

sense 'conceptual agency' does not exist, except as a form of shorthand. 'Conceptual agency' is an emergent artefact from the interpretation of data about people by researchers.

Concepts within the theoretical framework of that research were retrospective constructions by the researcher. However, there were other concepts involved within the research – those it was inferred that the participants did or did not possess. Teacher X had a concept of, say, energy conservation (which she suspected might be a misconception) and proceeded to teach energy topics in a way that minimised the risk of pupils asking her questions about it. In this sense, the concept possessed did have an influence on actions i.e. it did have some 'agency'.

This is relevant to the current research. As the research was being carried out, the focus was on what people thought, felt and did; it was on events and developments and people's perceptions of the factors influencing these. The focus was not directly on the concepts of policy, practice and professionalism. These concepts can be seen as part of the analytical equipment of the researcher. However, as in the example discussed above, there is one concept that can be seen as being in the ownership of the participants as well as the researcher and that is professionalism. Participants in the research were assumed to possess, in some form or another, the concept of academic professionalism. This can be seen as a kind of broad attitude with cognitive and value components, the possession of which predisposes the possessor to act in some ways rather than others. In this sense, the concept of professionalism does have an influence on actions.

In the light of this discussion, it seems desirable to look more carefully at the primary concepts chosen for this research. In what kinds of concrete particulars are the general abstract notions grounded? 'Assessment policy' covers policy as content (the content of policy documents, for example), policy as a process of policy-making (what the policy makers did and thought) and policy as a process of policy-activity (including what practitioners did within this contested terrain). 'Assessment practice' covers everything that assessment practitioners did in relation to the assessing of students and on their perceptions of why they did it and why their practice changed. 'Academic professionalism' was both a researcher construct for the theoretical framework and a concept assumed to be possessed in some form or another by the participants in the research, one which influenced their perceptions of both policy and practice. It was thus a purpose of the research to reveal (albeit indirectly) aspects of the concept of professionalism possessed by the staff studied (and to interpret these in terms of a stipulative definition of professionalism developed by the researcher) and then to consider the relationships that existed between professionalism and the other concepts. (There would also be a temptation to discuss the types of relationships that might be encouraged.)

(e) The issues imported into the research and those emerging from it have to be distinguished and discussed.

(What are to count as issues in this research?)

The word issue is now very common in educational discourse. Whereas in the past there were questions to be answered, or problems to be solved (or at least resolved), it sometimes seems, in our less confident age, that there are only 'issues to be addressed'. This is yet another word to which more than one meaning can be attached. For our present purposes, two of these are important. Firstly, an issue is something which emerges from something else (i.e. it is an outcome, an outflow, a discharge); secondly an issue is an important subject of debate about which there are clearly differing views (a matter at issue).

Within educational research itself, it is now common practice to conclude a report with 'emerging issues' (i.e. matters at issue which issue from the research). However, it is also possible to embark on research with issues in mind, one that the research should address. In this present research, the word will be used with both these meanings. Before the empirical work of the research, the researcher had issues in mind which the research was intended to explore (these will be labelled 'imported issues'); during and after the empirical work, analysis was guided in part by these imported issues, but it also revealed unpredicted issues important to the participants (these are labelled 'emerging issues').

This position is very similar to that of Stake (1995). Although he was dealing specifically with case study research, he made two important distinctions relevant to other types of research. Firstly, issues may be *etic*, in that they are initiated or brought in from the outside by the researcher or there may be *emic*, in that they are issues significant to the actors in the research. Secondly, there are intrinsic case studies in which the case studied is itself of primary importance and there are instrumental case studies where the research is conducted on a case to gain understanding of something else; in the former, the *emic* issues are the more important; in the latter the *etic* issues constitute the 'something else'. He then proceeds to make the point that what he calls 'issue questions' are the most productive way of ensuring that the research has an adequate conceptual basis or theoretical framework. This poses a very important question, not yet adequately answered in the methodology literature as far as this writer is aware, about how issues, research questions and underpinning concepts are related. As a contribution to an answer, the following is offered.

- (a) Research questions which are topical information questions have little conceptual significance, because they result in description rather than explanation. ('How are students on course X currently assessed?')**
- (b) When research questions are issue questions, it should be possible to detect significant concepts within them. ('What impact does changing assessment practice have on professionalism?')**
- (c) Much educational research is both intrinsic and instrumental; both *emic* and *etic* issues are likely to be involved.**

- (d) In educational research imported issues are sometimes abandoned; more commonly they are re-defined and thus become emerging issues.
- (e) In educational research, imported concepts may be abandoned; more commonly they are re-defined by clarifying the relationships between them.
- (f) *The research must be located in context.*

(What aspects of context are relevant to this research?)

If the most frequent bit of advice given to students is to ensure their research goes beyond the merely descriptive, then the second most frequent bit is probably to locate the study in a context. This too requires some unpacking; it means at least four different things.

Firstly, there is the **time context**. Here the student is advised to describe those relevant things that precede the research and perhaps also those they hope will follow it; their particular piece of work is not to be seen as isolated from past and future, independent of the particular time it was conducted. Secondly, there is the **place context**; this expects some relevant description of the country, institution, department, home (or whatever) in which the chosen research sites were located. This contextualisation (as with the previous type) is related to questions about the generalisability of research claims. For example, if readers of the research report are to be able to make any sound naturalistic generalisation to other settings, then they have to be helped to judge how similar to, and how different from, the location of the research are those other places which the research might illuminate. The third is the **policy context**. It is difficult to imagine any subject of social scientific research, certainly in education, which is not influenced in some way by statements about principles and courses of action from bodies with legitimate authority within the area of the topic. Likely sources of policy, i.e. the determinants of the policy context, are the government, the institution, a unit of organisation or a professional body. The fourth is the **theoretical context**; this is taken here to be the same as the conceptual basis or theoretical framework.

Reviewing the literature of theory and of policy will normally take care of most of the time context (i.e. that the relevant antecedents are revealed) although there could be simple events that are influential precursors of the research. Trying to indicate the 'future' context is much more problematic; it is this which normally results in some discussion of the possible impact and implications of the research, in proposals for dissemination and follow-up activity and in recommendations for future research in the area.

It is not necessary, or even appropriate, to deal with these four aspects of context together and in one section. They should, however, all be dealt with somewhere.

Final note

If anyone is still reading this, my congratulations and my apologies. I am acutely aware that this appendix is both over-long and much too short. There is a book here that I know now I wish to write. I also know that such a book needs much more and much better scholarship than supports this treatment if it is to satisfy me, let alone be seen as satisfactory by others.

REFERENCES

- Bliss, J., Monk, M. and Ogborn, J. (1983) *Qualitative Data Analysis for Educational Research*, London: Croom Helm.
- Blumer, H. (1969) *Symbolic Interactionism: Perspectives and Methods*, Englewood Cliffs NJ: Prentice-Hall
- Crotty, M. (1998) *The Foundations of Social Research: Meaning and Perspective in the Research Process*, London: Sage.
- Glaser, B.G. and Strauss, A.L. (1967) *The discovery of grounded theory: strategies for qualitative research*, Chicago: Aldine.
- Hammersley, M. (1985) 'Ethnography: What it is and what it offers', in *Research and Evaluation Methods in Special Education*, Hegarty, S and Evans, P. (eds), Philadelphia: Nefar-Nelson.
- Hammersley, M. (1996) *Educational Research in Action (The E835 Study Guide)*, Milton Keynes: The Open University.
- Hammersley, M. (2002) Personal correspondence.
- Harlen, W. and Holroyd, C. (1995) *Confidence and Understanding in Teaching Science in Primary Schools*, Edinburgh: Scottish Centre for Research in Education.
- Hempel, C.G. (1966) *Philosophy of Natural Science*, Englewood Cliffs NJ: Prentice-Hall.
- Huberman, A.M. and Miles, M.B. (1994) 'Data Management and Analysis Methods' in *Handbook of Qualitative Research*, Denzin, N.K. and Lincoln, Y.S. (eds), London: Sage.
- Koestler, A. (1964) *The Act of Creation*, London: Hutchinson.
- Nagel, E. (1961) *The Structure of Science: Problems in the logic of explanation*. New York: Harcourt Brace.

Stake, R.E. (1995) *The Art of Case Study Research*, London: Sage.

Strauss, A. and Corbin, J. (1994) 'Grounded Theory Methodology: An Overview' in *Handbook of Qualitative Research*, Denzin, N.K. and Lincoln, Y.S. (eds) London: Sage.

Weinberg, S. (1993) *Dreams of a Final Theory*, London: Hutchinson.

ASSESSMENT IN HIGHER EDUCATION

SUBJECT AREA I

Chemistry: a descriptive report

Version 2

A. SETTING THE SCENE

(a) The Research Purposes

The research aims to illuminate the underlying rationale for practice in assessing student learning; it aims for description and explanation, there is no evaluative purpose. This phase of the research is guided by three research questions: Why do we assess students in the ways we do? When assessment practice changes, why does it change? What issues in assessment emerge as important?

(b) The Courses

The focus was on assessment in chemistry at Level I and at Honours Level. The study was limited to cover six courses: two at Level 1 (Chemistry 1 and General Chemistry 1) and four at Honours Level (BSc Chemistry - 4H, BSc Chemistry with Medicinal Chemistry - CMC 4H, MSci Chemistry - 4M and MSci Chemistry with Medicinal Chemistry - CMC 4M).

The numbers of students on these courses in session 2000-2001 are as shown.

Course	No. of students
Chemistry 1	475
General Chemistry 1	181
Chemistry 4H	8
Chemistry with Medicinal Chemistry CMC 4H	19
Chemistry 4M	7
Chemistry with Medicinal Chemistry CMC 4M	14

(c) The People Interviewed

General approval was obtained from the head of department; he nominated a senior member of staff as the person for subsequent liaison. This member was asked to suggest the names of about eight members of academic staff whom it was recommended should be approached for interview. The request was for the names of people who were heavily involved in some way with the assessment of chemistry students at either Level 1 or Honours Level or both. It was requested that the people suggested should include the following: the chair of the teaching committee; the examinations officer; the class head for Level 1 chemistry; the class head(s) for Honours Level chemistry; an experienced lecturer with no current responsibility for the design or management of assessment; a lecturer recently appointed to this University, but with experience in assessment elsewhere; a recently appointed lecturer for whom the current post was their first. Nine names were provided.

These nine were approached by letter at the beginning of October; the letter covered the purposes of the research, the main topic areas for interview and an indication of the time that might be required; enclosed was a protocol making explicit the 'rules of engagement' and covering aspects of approval and anonymity. The letter was followed up by e-mail. Six people

immediately agreed and a date for interview was fixed; three people were unwilling to be interviewed within the suggested period. Two replacements were sought and obtained. Eight people were then interviewed in the last two weeks of October; these eight covered the functions/categories outlined above.

There are currently thirty-eight members of academic staff in the chemistry department; it is not claimed that the eight interviewed make up a representative sample of this population. Rather they form a purposive sample, chosen because they satisfy the criterion of being well-qualified, through a variety of types of experience, to speak on the topics in the interview. There is no intention to make any empirical generalisations beyond this group.

Seven interviews were recorded on audio-tape with the subjects' permission; the eighth person chose not to be taped. Interviews lasted on average for 46 minutes (excluding points in preamble and farewells) and ranged from 38 to 57 minutes. The experience of assessing students varied from 3 years to over 30. One person chose to focus on Level 1 only and three on Honours only; four talked about both levels. Data about Level 1 thus came from five people and about Honours from seven people.

After these interviews, five other people were contacted: two were suggested during the interviews ('It would be a good idea if you spoke to....'); three were chosen by the interviewer to check and/or amplify points which seemed unclear. Four short interviews were then conducted face-to-face and there was one e-mail interaction. The five included two lecturers in chemistry, the head of department, the dean of the science faculty (himself a chemist) and a member of staff of the Teaching and Learning Service (also a chemist).

(d) Chemistry as a Subject Area

In choosing chemistry as one of the subject areas to be explored, there is an assumption that it makes sense to treat chemistry as one subject area i.e. as one discipline, albeit containing specialisms within it. This assumption was tested in interview.

The external boundaries of chemistry, it was said, kept moving and were permeable; within chemistry the specialists re-grouped and re-named themselves, for example as physical, inorganic and organic chemists or as theory, synthesis and catalysis chemists. The people interviewed all saw themselves as 'chemists' and chemistry as one subject area.

I still think of myself as a chemist. A chemist is a chemist and distinct from a physicist or a biologist. Chemists have a way of looking at things which is uniquely chemical in some strange way. [1]

I see chemistry as one big subject and we teach and assess it as if it were one unified subject area. [2]

Intriguingly, although all eight people saw chemistry as one subject area, five of them thought that some of their colleagues did not.

I treat it as one big subject. Other people only see their own little world within it and would chop it up into bits. The people I think are the best chemists have a broad understanding across the whole subject; they talk a lot with people from the range of specialisms. We should all talk more with each other.... [3]

(e) 'The four quadrants' Where is chemistry?

Interviewees were given the diagram below with this introduction. 'It is said that different subjects are differently located on this diagram. The axes represent what is emphasised in assessment at honours level. Please do two things for me: (a) comment on the framework and (b) whatever your reservations, place chemistry where you think most appropriate.'

People tended to talk about what chemistry courses **should** be like, rather than what was actually emphasised in honours assessment. It may be that there is a genuinely close correspondence between what courses should be like and what actually gets assessed i.e. between the rhetoric of course purposes and the reality of assessment procedures. Confirmation of this would require not only an analysis of what is sought through the assessment instruments and procedures, but also some scrutiny of what gets rewarded in student responses. This goes well beyond the present study.

Several people said they did not see the poles of each dimension as antithetical; in addition, two said that they did not see the axes as genuinely distinct.

You have to possess a degree which is academically sound to be employable; you have to understand the concepts before you can evaluate, say, alternative routes in chemical synthesis. These things are not mutually exclusive. [4]

You have not just to understand concepts, but to apply them in problem-solving; chemistry is a problem-solving subject and problem-solving is a transferable skill; therefore understanding and fitness to practise are not separate. [2]

One person was unwilling to locate chemistry in the framework. The average positioning of the other seven was in the lower-left quadrant (at -3, -1); within this group, six people located chemistry within this quadrant, the seventh placed chemistry in the upper-left quadrant on the

grounds that the honours chemistry course was accepted by the Royal Society of Chemistry for membership i.e. fitted graduates to practise chemistry.

Notwithstanding these reservations, it can be claimed that honours chemistry was thought to put a significant emphasis on assessing understanding of a set of key concepts and to put (slightly) more emphasis on the 'pure' rather than the 'applied' assessment emphasis. Chemistry was seen as being located in the 'hard and pure' quadrant, but in a less decisive way than the researcher had hypothesised.

B. STUDENT ASSESSMENT: THE PRESENT

(a) How are students on these courses currently assessed?

The descriptions which follow are summary versions of those which appear in course documentation for session 2000 – 2001.

Level 1 courses, Chemistry and General Chemistry, are assessed in the same way.

Method of assessment	Weighting for final grade
Laboratory work (report assessed)	10%
Class exam (2 hour paper – work of term1)	30%
Short tests (4 tests in lecture time)	10%
End of course exam (3 hour paper)	50%

The weight attributable to assessment other than the end-of-course examination is thus 50%.

Honours course (BSc) Chemistry 4H is assessed as below.

Method of assessment	Weighting for final result
Carry-over mark from third year	10%
Research project:	
work during lab	40%
the thesis	40%
oral on thesis	20%
Degree exam (4 x 3 hour papers)	18% (20% of 90%)
	72% (80% of 90%)

Assessment in Chemistry with Medicinal Chemistry – CMC 4H is very similar, the only difference being that one degree paper is replaced by two half-papers. The weight attributable to assessment other than the final examinations, in both courses, is thus 28%.

The MSci Honours courses, Chemistry 4M and CMC 4M, are assessed as below.

Method of assessment	Weighting for final result (approx)
Carry-over from year 3 exam	7%
Carry-over from year 3 essay	4%
Carry-over from year 3 project essay	4%
Research project	13%
Poster and oral presentation	5%
Degree exam (4 x 3 hour papers as 4H)	54%
Additional MSci paper (1 x 3 hours)	13%

In the present MSci courses, the weighting attributable to assessment other than the final examinations is thus 33%.

The University also now offers a (4 + 1) year MSci course; in this either a work or European placement occurs between years 3 and 4. Students will start these placements for the first time in September 2000. For the European placement basic competence in the appropriate language is a requirement and is assessed; this is a base-line prerequisite i.e. the level of competence does not affect the final degree awarded. For both types of placement there is formal assessment of a report and presentation during the placement and of a thesis and presentation after the placement; these carry equal weight and have a significant influence in arriving at the final degree award. The report and presentation are assessed jointly by an academic supervisor and the placement supervisor; the thesis and presentation are assessed by a departmental panel.

For this new (4 + 1) MSci, the above assessment summary table has to be modified to the following.

Method of assessment	Weighting for final award (approx)
Carry-over from year 3 exam	5%
Carry over from year 3 essay	3%
Carry-over from year 3 project essay	3%
Research project	11%
Poster and oral presentation	4%
Degree exam (4 x 3 hour papers)	43%
Additional MSci paper	11%
Report/presentation – during	10%
Thesis/presentation – after	10%

In the (4 + 1) MSci, the weight attributable to assessment other than the final degree examinations is 46%. Such students are not excused any final degree papers; the weight attributed to these appears less only because these students have more assessments overall.

As we move from BSc to MSci to extended MSci, the relative weighting given to forms of continuous assessment increases, from 28% to 33% to 46%. The amount of assessment by final degree examinations does not go down; it only forms a decreasing percentage of a greater whole. Final 'big-bang' examinations remain big; when perceived as less important than they were, this is because continuous summative assessment has been added in to the assessment scheme.

B (b) What is seen as good? Level 1

Three things were seen as pleasing at this level: the effectiveness of the assessment scheme as a whole, the four short tests and the opportunities for informal diagnostic assessment.

The courses and their assessment procedures were said to work 'extremely well' as judged by student feedback, pass-rates and the proportion of students going on to study chemistry at higher levels. Students on the whole appeared 'to enjoy, to understand and to continue'. One factor contributing to the success of the assessment scheme was thought to be the clarity and quality of the documentation describing it to students.

The four short tests were judged a very successful innovation in assessment. They motivated students to work throughout the year and allowed the detection of those who were coping less

well. Because the tests were related to a preparatory workshop, students tended to score highly and they found the experience of success positively reinforcing.

Workshops, and the laboratory work throughout the year, provided good opportunities to spot those whose understanding was not developing as well as it should. Although laboratory work, as described in reports of the experimental work undertaken counted for relatively little in a summative sense (10% overall) it was seen as providing many opportunities for formative and diagnostic assessment and for that informal assessment which is integral to effective teaching.

It was very noticeable that most people responded by saying 'I think that X is good, but I have a worry about....' And in the worries expressed there were clear differences of opinion. For example, one person said that the approach was very traditional in that 'we still rely a great deal on conventional examinations'; another said that s/he was not personally keen on giving as much weight (30%) to the mid-course examination because any assessment scheme should give more weight to the understanding developed towards the end of a course. Whereas the four short tests were generally favoured by all, two people had some reservations: the tests were not sufficiently 'rigorous' and 'students score too highly on them because we spoon-feed them'. Although laboratories were the ideal place to remedy deficiencies in knowledge and understanding, the people who staffed laboratories varied very widely in their ability to capitalise on opportunities provided. One person said that assessment obviously worked well because pass-rates were high; another said that the low pass-rate in General Chemistry was 'endlessly discussed'.

What is seen as good? Honours Level

There was a high level of satisfaction with how students were assessed in the four BSc and MSci courses considered.

We have an assessment scheme which makes sense. We are examining the right things by the right methods; the conditions are right and the procedures work smoothly. [4]

We are pretty happy with what we are doing and we should stick to it. [1]

The assessment scheme for MSci is essentially that for BSc with some extras added on. There were four reasons given for satisfaction with the core assessment common to both: the quality of the final degree papers; the balance and timing of the main assessment components; the student abilities developed and revealed by the project; the evidence from external sources.

The final examination papers were regularly praised highly for their quality. This was seen as attributable to two sources. Firstly, questions were carefully set.

We spend a hell of a time on setting, scrutiny, careful checking and modification of questions to see they reflect properly the course coverage and are very clearly worded. [5]

Secondly, these papers were thought to have a pleasingly high problem-solving emphasis; some questions did not stem directly from any one set of lectures but required students to bring together their understanding from several sources and to apply it to an unfamiliar situation. This might be done, for example, by setting a problem-solving question based on a synthesis mechanism reported in a recent journal article.

Although there was still a significant weight attached to the final examinations (up to 72% of the total) this was seen as right and appropriate, in that it allows cross-referencing, cumulation and consolidation of understanding, the pulling together (or pooling?) of knowledge from different areas. The carry-over mark from the third year (about which two people had reservations) was seen by most as freeing the final exams to put less emphasis on the recall of knowledge and the checking of basic understanding and to put more emphasis on higher cognitive abilities. The project although it might be thought to have relatively little importance (18%) carried sufficient weight for students to take it seriously.

The research project was consistently praised for developing skills appropriate to chemistry and of real value in later employment: practical laboratory skills, report-writing and oral communicative competence. Students were said usually to derive both benefit and pleasure from pursuing a topic to some depth and with a measure of independence. It was also pointed out that for staff, supervision of projects meant both informal assessment of progress (with 'remedial' teaching as required) and more formal assessment. The involvement of at least three different people in the project assessment meant that staff could be 'pretty confident that students get the marks they deserve'.

Several staff, possibly aware of a risk that they sounded too easily satisfied, emphasised that there was sound evidence that assessment methods worked well from three distinct external sources: external examiners (who were said to view these assessment procedures more positively than the interviewees who were themselves external examiners viewed assessment elsewhere), the Royal Society of Chemistry (which had approved the level of problem-solving assessment) and employers (who trusted the class of honours awarded to give a valid indication of a graduate's ability).

Views on the extra elements in MSci assessment came from three people. There was approval for the carry-over from the year 3 essays, which dealt with new developments from the frontiers of research and with the planning and literature search for the project. The additional paper for MSci gave an opportunity for more student choice in doing work at greater depth. The posters and oral presentations were said to work extremely well; they encouraged useful skills for communicating with the scientific community i.e. being succinct and using helpful diagrams. Although their summative contribution was modest, it was enough for most students to take these aspects of the course seriously. The assessment of posters and oral presentations had required assessment criteria to be made more explicit than was usual in more conventional methods; it was also recognised by one person that careful explication alone was inadequate – different assessors had to work harder at gaining some shared understanding of what the criteria actually meant in practice.

C. STUDENT ASSESSMENT: LOOKING BACK

(a) Introduction

Interviewees were asked to identify what they saw as significant changes in how students have been assessed. It was decided beforehand not to restrict people to some arbitrary period of time, but to encourage them to reflect on the whole period over which their experience extended, however short or long that had been. This led to considerable variety of types of response. For example, one person said, 'I have only been here for a few years and there has been no dramatic change in assessment' and another said 'In the distant past we attached more importance to assessing practical work, then we retreated from that and then we began to change back again. Is that two changes or one?'

Interviewees were then asked why they thought a change had occurred. It was anticipated that this would be a difficult question to answer and so it proved. Change usually has many subtle

and interacting causes; it is possible to be involved in change and not to know why it is happening; it is possible that one knew why at the time but cannot now recall the reasons. However, no apology is necessary for asking the question; people's perceptions of reasons are arguably as important as the 'real' reasons, even if it could be agreed that these are ever accurately knowable.

Some clarification of terms used by the interviewees and used in this report is necessary here. Firstly, some staff referred to continuous assessment and some to continual assessment. Although continual (in the sense of frequently recurring) is perhaps more accurate than continuous (connected throughout in time), conventional current usage favours continuous. In what follows the term continuous assessment is used. Secondly continuous is taken by the writer to be the opposite of terminal: continuous assessment takes place during a course; terminal assessment occurs at the end of the course. These terms have to be very clearly distinguished from formative and summative. Formative assessment is taken to be assessment which provides the student with some kind of feedback intended to influence their future performance; summative assessment is assessment which provides grades or marks which are taken into account in making end-of-course judgements of success and failure (or of levels of success). Formative assessment is usually continuous, but continuous assessment may be summative as well as formative. Terminal assessment can have no formative impact during the course, but could well have a formative influence on subsequent, after-the-course learning.

(b) Significant Changes at Level 1

Staff identified four changes that they saw as significant: increased continuous, summative assessment; the introduction of the four short tests; more emphasis on formative assessment and subsequent learning support and more concern with the productivity of practical work. These are not all distinct; there are links.

Continuous summative assessment was said to have increased. In the past there had been two class exams which counted towards the award of a class-ticket ('duly performed the work of the class') which made the holder eligible to sit degree exams; marks awarded in the class exams did not count towards the final grade awarded, everything depended on the terminal degree examinations. Lab work, short tests and a single class exam now make up the continuous assessment which accounts, summatively, for 50% of the final grade.

The four short tests motivated students to work more consistently throughout the year and allowed them better to monitor their own progress and to receive early warnings of their failure to understand. The nature of the tests was such that the majority of students scored highly and felt that their efforts were being positively rewarded.

The continuous assessment as a whole was said to serve a useful formative function. Apart from what the results told students themselves directly, they allowed staff to identify those who were struggling and then to provide some support and remedial help. For example, the fifty lowest performing students in the tests were invited to attend additional tutorials. The regular submission of lab reports to on-the-spot tutor/demonstrators also provided feedback to students on their progress in lab-work and staff were on hand to assist with difficulties.

Changes in the assessment of practical work were thought significant but staff found some difficulty in describing them. People had the impression that its weight had increased, but could not recall what it had been and how great the increase was. The current modest weighting to practical work (10%) was thought to be about right; although students might wish more importance attached to it, it was said that staff would oppose any further increase in weighting. One person argued at some length that the assessment of practical work was

very subjective; tutor-demonstrators tended to be inconsistent in the application of assessment criteria; they tended to mark too highly and within a very restricted range and thus performance in practical work was not a good discriminator of different ability levels. However, others did say that considerable effort had been invested in clarifying the purposes of practical work, to introducing pre-laboratory preparation sessions and to making assessment criteria explicit. This strikes this writer as a good example of an assessment method which is perceived to have high face validity (or curriculum fidelity) but is distrusted on grounds of low reliability; it also seems to reveal a tension between criterion-referenced thinking (we are testing for mastery of skills) and norm-referenced thinking (we are testing to spread people out).

(c) Significant Changes at Honours Level

Assessment at Honours Level within BSc courses was said not to have changed much in recent years: 'We haven't tinkered with it much, because it seems to work well'; 'If it's not broke, don't fix it'. Nevertheless three changes were thought important by several people: the increased weight given to the project; the introduction of a carry-over mark from Junior Honours in year 3; a 'tightening-up' of the assessment of laboratory work (in year 3).

The honours research project now carries the same weight as each of the four final degree papers. Assessment applies to the work during the lab, the written report of the project and an oral examination on that report. The project was seen as being an effective assessment method and one carrying appropriate weight: students spent a lot of time on the project and this was now recognised in assessment; they developed rather different skills to those assessed in formal written papers and these skills were relevant to further work in chemical research and to future employment. The main justifications for this assessment thus lay in curriculum fidelity and its contribution to transferable skills.

Previously success in Junior Honours was a condition for entry to Senior Honours, but had no influence on the final outcome. In the carry-over mark from Year 3 there was now an element of continuous summative assessment. No-one wished to see this gain greater significance; the weighting was seen as adequate to encourage students to take it seriously. Again chemistry was seen as a cumulative subject of which students could only gain deep understanding towards the conclusion of the course; it was right then that end-of-course assessment was given most weight. (The researcher was instructed not to use the hideous term 'exit velocity'.)

Although it had minimal direct impact on summative assessment for the final honours award, three people said that changes in the assessment of practical work in year 3 were significant. One person was keenly interested in this 'tightening-up'; what s/he said raises fascinating questions about assessment and will be quoted at some length without comment.

Interviewee [6] *.....so there was a large number of students and when they came up to the people marking them they put a bit of pressure on them. The labs hadn't been reformed for a long time. The same labs had been running for years and years. And almost all the students were getting nine or ten out of ten. We weren't happy with that and we have tightened up in the way we require them to submit their work. There was a tendency for things not to be written up until the last moment and that put tremendous pressure on staff and allowed real possibilities of plagiarism. And we had people performing to a level that their written work indicated they weren't capable of. So we've addressed these problems by saying that they can't do more than two experiments before they have them marked and are allowed to continue and by distributing new guidelines to markers... We now say what type of performance is expected to earn such and such a level of mark and this works much better. Sorry, I've gone on a bit and rather lost the place....*

Researcher *No, you haven't. Students have to get reports marked after every two experiments and the assessors work to a description of what to look for, for each grade?*

Interviewee *That's it. Marking things at an even pace. And if you just do the minimum in completing the work and writing it up you'll just get a five. And if a marker awards a 10, then I'll want to know why. What is so very distinguished about it?*

Researcher *Improving assessment of lab work is the most significant change?*

Interviewee *Definitely.*

Researcher *And the reason for the change was...?*

Interviewee *Hopefully to overcome the very irritating impression that some students were just not good enough for the kinds of marks they were being given. We felt students were being given marks they hadn't really earned; that their learning wasn't really as good as we seemed to be assessing it to be. And also to even out the burden of work for both students and staff.*

Four other changes were mentioned each by only one person: more opportunities to exercise communication skills, although these were not, in BSc, formally assessed; the deletion of an essay paper from degree examinations, because students did it so badly; the separation of second-class honours into two divisions, which was said to have led to more formulaic aggregation procedures and the introduction of designated degrees in which marks for individual papers, rather than courses, were aggregated in a 'somewhat bureaucratic way which paid scant regard to educational coherence'.

The advent of MSci courses had resulted in a considerable diversification in assessment methods. (See B(a)) Significant changes were the introduction of essays, and of oral and poster presentations, all of which were continuous (and gave more opportunities for constructive feedback) with a summative weighting. The four-plus one MSci would mean the further addition, to cover the placement year, of a report and presentation during the year and a thesis and presentation after it. For European placements language competence would be required (and assessed) but not count in the final degree award; placement supervisors would be involved as assessors. This last was the first example, in chemistry, of assessment being carried out by some one other than a lecturer/tutor/demonstrator within the department.

(d) Reasons for these Changes.

At the outset there was a hope that people would identify the change they saw as most significant and then identify the reasons why that change had occurred. Either this hope was naïve and misconceived (given the nature of causality in human affairs) or the interviewing technique was inadequate to the task. What happened in the majority of cases was that two or three significant changes were described and two or three influencing factors were then identified; each factor was perceived as applying to each change to some greater or lesser extent. It was not possible to relate the potency of any one factor to one specific change.

There were three broad types of change described. Firstly there were relatively long-term evolutionary trends; of these the one most frequently mentioned was the increasing weight attached to continuous summative assessment. Secondly, there were specific changes in

assessment methods introduced within existing courses, what might be described as innovative events; the introduction of the four short tests to Level 1 was an example of such innovation. Thirdly, there were changes in assessment practice which were components of a larger change in course structure, either the creation of a new course or some radical revision of an existing one. When it was necessary to think about all aspects of course design, it was inevitable that assessment would be considered and it was likely that significant changes would be introduced. The creation of the four-year MSci course and its later extension to a four-plus-one structure are examples in this third category.

Several reasons were given for these changes. They fell under three headings: learning enhancement, conformity to fashion and expediency. Learning enhancement in some form or another was mentioned much more frequently, but it was noticeable that most people tended to temper one 'worthy' student-centre factor with some other factor they themselves described as 'less educationally respectable'. The reasons classified as learning enhancement were to do with the quantity of student learning, its quality and its appropriateness (to the nature of chemistry as a subject).

..... because we wanted them to work consistently harder throughout the course. [7]

..... to give us a chance of diagnosing poor understanding and doing something about it. [2]

..... because chemistry is a practical, problem-solving subject, our assessment changed to reflect that better. [4]

Conformity to fashion was a frequent reason given both for evolutionary trends in assessment and for assessment change integral to course design.

It was all part of the fashion favoured by people in high places. [extra]

Everybody else seemed to be talking about transferable skills, so we thought we had better do it as well. [8]

It just seemed to fit in with the climate of the times. [5]

No doubt there are situations where it is expedient to conform with fashion, but the reasons categorised as expedient seem rather different from 'mere' fashion. They include things like the following.

We did that just to get better pass-rates. [7]

I think we did that mainly just to get the students on our side. [1]

If we hadn't introduced this kind of MSci, we'd have lost out in competition with other universities. [6]

It intrigued this researcher that, apart perhaps from a few references to 'people in high places' and to 'those up the hill', there was no allusion of any kind to assessment change being required by extra-departmental policy. The only explicit references to faculty or university policy came in describing changes in assessment practice that had **not** occurred; these were changes that would have been necessary if a policy of modularisation had not been

successfully resisted. One quotation will serve to illustrate points made by several members of staff.

There was a university and faculty policy for modularisation. We had this attack on us and we successfully defended ourselves. It's the nature of the subject. Modularisation has been a disaster in some chemistry courses, like A-level. Students have come to me and said, 'I don't know as much as I should – once we'd done a module we just sort of forgot it'. I'm very much against this sort of thing – it just doesn't make sense in chemistry when everything you do builds on what went before. If we had had to modularise chemistry into short, fat courses we'd have had to change the assessment procedures and the modules would have ended up with premature final assessment and over-assessment both from staff and student points of view. [2]

(e) Assessment Change: Exploring One Example

It was claimed above that the introduction of 'the four short tests' to Chemistry 1 was a discrete innovation in assessment to be explained by a wish to motivate students to work more consistently throughout the course. This is credible but too simple. Additional evidence was uncovered which shows the limitations of both the descriptive and explanatory components of such a claim.

In the early 1990s the Ordinary Chemistry class was very big and catered for a very diverse student population. Students varied widely in their intended degree paths; for example, some wished to achieve honours in chemistry and some needed a basic qualification in chemistry to study biology. Students also varied widely in their previous experience of, and qualifications in, chemistry; the latter varied from nothing to A-level. It was decided to construct a new course, one which came to be the General Chemistry course listed in section B(a). This new course was planned by a team of eight staff and their planning was guided by ten educational principles, derived in the main from research in science education. (The Science Education Unit was based in the chemistry department; the director of the unit was an influential figure with an international reputation in science education research.) Of these ten principles ('The Ten Commandments' – see ¹Gray), two were directly relevant to assessment practice.

At that time Ordinary Chemistry 1, like other course in the science faculty, had two class exams. These conferred eligibility to sit the degree exam, but did not count towards the level of the final award; students scoring highly in these two exams could gain exemption from the degree exam. The General Chemistry team decided that there should be no exemptions in their course and that their two class exams would have a summative weighting. In the overall scheme, lab-work would contribute 10% to the final result, the two class exams would each count for 20% and the degree exam 50%. Also included in the assessment scheme would be short diagnostic tests with formative but not summative significance; students would be strongly encouraged to sit these ('through enlightened self-interest') but sitting them would remain entirely voluntary. The reasons given for this scheme were (i) to motivate students to work consistently and to take some responsibility for their own learning, (ii) to reduce the seriousness of terminal examinations and (iii) to be consistent with assessment processes experienced in schools. This scheme operated for two sessions, 1993/94 and 1994/95.

Class exams in different science subjects were held at different times in the session. As one part of the faculty and university policy for modularisation, the timing of end of module examinations was brought into line – these were all to be held in week 13 (end of January).

¹ Gray, C. (1997) *A study of factors affecting a curriculum innovation in University Chemistry*, Unpublished PhD Thesis, University of Glasgow.

Even when a subject resisted full-scale modularisation, as did chemistry, this meant that in any year-long (double module) course, one summative exam had to be held in January.

The leaders of Chemistry 1 then decided that chemistry students needed an earlier assessment than in January both to encourage them to work and to give them earlier feedback on their progress, or lack of it. General Chemistry provided an example of the use of short tests which was then adopted by Chemistry 1, but with a difference: these were now made compulsory and given a summative role. This decision was then adopted by General Chemistry. Thus, from session 1995/96, both Chemistry 1 and General Chemistry had the assessment scheme which was summarised in Section B(a).

The introduction of 'the four short tests' to Chemistry 1 now appears not simply as a discrete innovation within an existing course. An assessment change which was a component of the overall process of course design for a new course (General Chemistry), became an influence making an innovation more likely in an existing course (Chemistry 1) and these two changes can be construed as steps in one general evolutionary trend. The three 'types' of change suggested in C(d) are inter-related.

One important factor influencing the introduction of the short tests was a general view amongst teaching staff that student learning could be improved by an assessment change. In the above account it can be seen that many other factors were operating. These included the indirect influence of educational research, the impact of one strong personality, pressures from increased number of and diversity within students, acceptance of and resistance to faculty and institutional policy and awareness of relevant assessment changes in other courses. The views of interviewees on these, and other, factors are considered in Section E.

D. STUDENT ASSESSMENT: LOOKING FORWARD

(a) Changes Wished

Interviewees were asked what changes they personally would like to see in how students were assessed, assuming that the conditions for introducing such changes could be made favourable.

Level 1

Two people did not identify any desired changes, because 'assessment is currently very satisfactory'. Four recommendations for change were given by three other people. There should be less choice within examination papers; this required a definition of that core of chemical knowledge which students really **ought** to understand – at present it was too easy to miss out chunks of important content. Secondly, there should be a departmental effort to work out what could be appropriately assessed in chemistry by objective testing; in the longer run this would save staff time on marking. Thirdly, there was a need for more systematic and rigorous assessment of lab-work to increase the standard of students' practical skills. Thirdly, because marking is parcelled out amongst so many, there was a need for greater departmental effort in explicating and achieving a common understanding of assessment criteria. It may or may not be significant that the most strongly held views for change came from the least experienced members of staff.

Those who followed up the prompt on favourable conditions talked of a greater recognition, within workload calculations, of the time that had to be devoted to marking for it to be consistent and for the spreading out of the marking load away from the most hectic periods. One, however, was very clear that assessment itself did not need more time; what was

crucially required was more staff time to support the students identified by assessment as not doing well.

Honours Level

Four people did not wish to see any change, again because 'we do assessment well'. One person extended to this level the argument for defining core content so that students could not opt out of some content areas (said to be especially common within Chemistry with Medicinal Chemistry). Another wished to see the regular testing of student competence to perform the types of chemical calculation routinely useful in chemical industry; at present these were only touched on and the skills were regularly lost.

The remaining member of staff identified six desired changes: (i) more weight to be attached to the project oral – 'oral communication really matters'; (ii) within the BSc there should be the same importance attached to communication skills in general as within MSci; (iii) the value of essay assessment in MSci should be reviewed and one or more essays dropped; (iv) structured problem-solving should be assessed continuously, not just under exam conditions; (v) more imaginative strategies of detecting and circumventing plagiarism need to be introduced and (vi) more systematic assessment of practical work should take place against more explicit criteria. When this person was asked if these ideas were often discussed, the reply was that they were ventilated in the coffee room but 'I have no idea what gets discussed in the Teaching Committee'.

(b) Changes that may be Required

Interviewees were asked what future changes in assessment they thought might be required of them, whether they welcomed such changes or not. Staff did not speak at length. Either they saw little to fear or saw no point in speculation. ('Not really – I don't feel anything is about to be forced upon us' and 'I don't know what's coming up and I'm not terribly interested'.)

When people did foresee changes, they saw them as being required by extra-departmental pressure; perhaps the form of the question made it unlikely that people would predict that departmental thinking would require them as individuals to change. It was, however, very noticeable that in clear distinction to when reflecting about past changes, people did talk regularly about the future impact of **policy**. Such policy requirements were seen as coming either from some national, supra-institutional source (such as QAA) or from some institutional, extra-departmental source – whether this latter was The University (i.e. central, senior management) or the faculty (of science) was vague.

Of changes that might be required, only one was viewed in a definitely positive light: the University or faculty might require the department to do more to detect and to combat plagiarism. There was an ambivalence about changes that might be required by national policy; several staff made it clear that moves from QAA towards national standardisation of courses, both in their content and assessment, would be unwelcome; on the other hand national policy requirements which could be perceived as being channelled through, or endorsed by, a subject-based organisation might not be resisted. One person illustrated this as follows.

Benchmarking probably won't affect us much. The RSC is involved and Prof X [a member of the department] has been on the benchmarking group and we've been consulted. We are probably there already. I do approve of a core curriculum, that's reasonable enough, but there won't be a prescription on assessment – agreement on what's ideal is pretty unlikely. [1]

There were unwelcome things each receiving only one mention: further modularisation (as university policy) encouraging premature and over-assessment; university policy requiring even more continuous summative assessment; university policy attempting more uniformity in assessment reporting, seen as likely to 'waste yet more paper'; increased numbers of students requiring economies of time through more objective testing.

If there was one thread running through most predictions it was that external policy requirements always seem to be in the direction of intensifying the current 'audit culture' within higher education.

..... this would be yet another development of the audit culture which always requires you to spend more time documenting what you do, rather than doing it or doing it better. Attention will shift from assessing students well to proving to others that we do it well; and this is an influence against doing it better. [6]

E. ASSESSMENT: FACTORS INFLUENCING CHANGE

In Section C(d) there is a summary of the reasons that interviewees identified as being important in past assessment changes. These were learning enhancement, conformity to fashion and practical expediency. There was little or no mention of policy except in so far as university policy required some form of modularisation of courses (which then required some re-consideration of assessment). In Section D(d) it was noted that 'policy' was seen as more likely to require assessment change in future.

Interviewees were presented with a list of factors possibly influencing assessment change. They were asked (a) to give their views of the importance of these factors in chemistry in the past and (b) to predict which would become more influential in future. The results are summarised below.

Factors influencing past change in assessment (Items given in decreasing order of importance)

Factors for which the most frequent response was 'very important for us'
General concern for the quality of student learning
External professional body (The Royal Society of Chemistry)

Factors for which the most frequent response was 'of some importance for us'
Change in course structure
Concern for transferable skills
External examiner recommended change
Increased diversity in student population
National policy required change
Awareness of assessment trends elsewhere

Factors for which the most frequent response was 'not important for us'
Institutional policy required change
Faculty policy required change
Availability of assessment technology
Students pressed for change
Increased number of students
One enthusiast persuaded us
Impact of educational/assessment theory

Factors expected to be more influential in future

National policy
Institutional policy
Faculty policy
The external professional body (RSC)
Availability of assessment technology

The above are summaries of the boxes ticked. Interviewees regularly commented on their choice of response. The main points about each factor are given below.

1. General concern for student learning was said to have been paramount. Changes in assessment had been introduced to influence how hard students worked and how much they learned. Interviewees did stress the quantity of learning rather than its quality. There was no overt recognition of the potential of appropriate assessment to encourage higher cognitive abilities or 'deep' rather than 'surface' learning.

2. The Royal Society of Chemistry was perceived to be of primary importance. It had not, however, acted as a change agent; rather any change initiated by the department had to be acceptable to the RSC.

The RSC hasn't actually done much about assessment yet; but, if they did, we would inevitably comply. It is very important that our degrees are recognised by the RSC; if they said we would not be validated unless we assessed in some way or another, then we would do it. [7]

3. There was general agreement that when a new course had been introduced, or there had been major change in an existing course, then it had been important to review assessment procedures as part of course planning.

4. In BSc courses increasing attention had been paid to transferable skills, although formal assessment of these had not been introduced. In MSci courses, a concern for the development of transferable skills had clearly been translated into assessment requirements. Interviewees fell into two groups: the majority saw attention to transferable skills as a sincere response to a legitimate concern; a few felt there had been cynical expediency.

We really consider that all the time – like heck we do. The only time we have gone beyond lip-service and window-dressing was in level 2. [Interactive Teaching Units] [5]

5. Although the regular 'tick' response was that external examiners had been of importance in influencing past changes, the spoken comments were illuminating. External examiners were not important at Level 1 ('They just leave us to get on with it...'); they were very important in helping to make assessment decisions about individual candidates – but not at all important in deciding how students as a whole should be assessed. Six of the eight interviewees provided very similar responses; we listen to what they say about assessment; we seriously consider it; we have always found good arguments to resist their recommendations. Changes in assessment practice sought by the department would need the approval of external examiners; any changes recommended to the department by external examiners were unlikely to be accepted.

6. Increased diversity within the student population had had little direct impact on how students were assessed. The indirect impact was, however, significant. The response to increased diversity at both Level 1 and Honours had been the creation of new courses; in creating a new course, more thought had been given to assessment procedures.

Two people did say that diversity was something that would have to be addressed. The department was having to deal with more and more weak students; better assessment was needed to identify inadequate learning earlier and more precisely.

7. National policy was potentially of importance, but had not so far seriously impacted on assessment, except possibly that Teaching Quality Assessment had encouraged thought about what existing assessment schemes might look like under external scrutiny. One person observed that one (minor) change had been brought in only because it was felt it would look good in the documentation.

8. Awareness of assessment trends elsewhere had been 'factored in somewhere or other'. Change had not been led by such considerations, but 'it is always reassuring to know that other chemistry departments are doing similar things'.

9. Institutional policy was not seen as having had a very strong influence, except in so far as it was the channel through which national requirements (e.g. for TQA) were made obvious. Two people asked the interviewer if there was any institutional policy on assessment, because 'I've never heard of such a thing'. (Up to now there have been Examination Regulations, which the University is currently working on to convert into a Code of Assessment.)

10. Faculty policy had not required change; things 'worked the other way around'. Changes initiated within the department had to be acceptable to the faculty. Faculty was seen as having been reactive rather than pro-active. It might be that Faculty would become more 'directively reactive' in future; it was seen as possible to become reactive in more creative ways.

11. People reported that there had been no significant changes in assessment practice because appropriate assessment technology had become available. Opinions were offered that it might well be useful in dealing more economically with large numbers at Level 1 or in formative assessment. Using computer technology at Honours Level and for summative assessment were viewed with considerable misgivings.

12. Students had never pressed for assessment change. If they had asked for it, it would have been considered but 'students do ask for very different things from one session to the next'.

13. Increased numbers of students had not directly brought about changes in assessment nor should it have done.

14. Except in so far as any projected change needs a champion, change had not come about because of the influence of one enthusiast. The department tended to think 'more in terms of general views than individual ones'.

15. Staff were asked if educational or assessment theory had been an important factor in past change. The responses to this were not unexpected: 'we don't have time to read books on assessment'; 'theory gives ideas which are impractical when you're faced with the reality of large classes'; 'theory doesn't influence practice directly but it does gradually affect the climate'. What were unexpected were those responses which went on to deal with the influence not of theory, but of a theorist. There had been in the past an assessment theorist

within the chemistry department who was generally agreed to have had a powerful influence on thinking about assessment; whether that influence had been generally benign or malign was a matter of profound disagreement.

Interviewees identified five factors which they expected to be more influential in future.

- National policy
- Institutional policy
- Faculty policy
- The external professional body (RSC)
- Availability of assessment technology

It is important to note that policy, whatever its source, was expected to become more significant.

F. DEPARTMENTAL CONCERNS

In this section four concerns are identified: the burden of assessment, plagiarism, objective testing and examination stress. The researcher had no preconceived idea that these would emerge as important to the people interviewed, questions about them were not included in the original interview schedule. In labelling these ‘concerns’, the word is used to mean a matter of interest where there is an undertone of worry or anxiety; conventionally in research reports these matters would be called emerging issues i.e. important subjects of debate. The writer has little evidence that these topics have been the subjects of debate so far; there is a temptation to recommend that this is what they should become.

(a) The Burden of Assessment

Every interviewee had something to say about the time they had to devote to assessment. The focus for concern was variously the time necessary to construct new and appropriately demanding questions for examination papers every year, the time-consuming aspect of devising mark schemes and model answers and, most importantly, the time required to mark, and double mark, essays and examination answers. In addition, it was regularly observed that any method of assessment which acquired a summative weighting it had not previously possessed took considerable extra time; whenever assessment started to matter more to students, it mattered more to staff ‘to do it properly’.

A remarkable feature was that although staff talked a lot about the overall assessment workload, they did **not** complain about it. Characteristic views were as follows.

Assessment is time-consuming but not too time-consuming. For most people it is just part of the job. I myself find marking wearisome and dispiriting, but it is just something that has to be done – a necessary evil. [6]

The marking gets spread around and gets done relatively quickly; nobody has ever complained much. I just shut myself away and do it. [1]

This lack of complaint might suggest that the assessment burden is not really a concern at all. However, three additional points should be taken into account: it was predicted by some that assessment workload would soon become a serious problem (due to increased numbers of students and the increasing proportion of those who were seen as critical and potentially litigious customers, to an increased diversity of assessment methods deployed and to a squeezing of teaching/assessment time by the research commitments of academics); a few

said that although they did not personally complain, some of their colleagues increasingly did; the majority said that although the overall workload was not excessive, there were peak times of the session when the load became 'well-nigh intolerable'.

Some people hate the work associated with assessment; they would be delighted if a less-time consuming scheme could be devised. Find a way to reduce the time spent on marking and everyone will burst into applause. [7]

(b) Plagiarism

Six of the eight people interviewed were concerned about plagiarism in some form or another. It was recognised that there probably had always been a minority of students guilty of some dishonesty or cheating in relation to assessment and that the present situation was far from a crisis; panic measures were certainly not advisable. Nevertheless, there was a clear view that plagiarism was on the increase and this was attributed to two causes. Firstly there was increased temptation for students to take short-cuts to success when more and more of them were undertaking part-time work to avoid financial problems; secondly, students were inevitably affected by the perceived decline in standards of honesty and probity in wider contemporary society.

'Newer' forms of assessment provided more opportunities for student dishonesty than did traditional examinations. Given the usual safeguards for these, assessors could be reasonably sure that answers produced under examination conditions were the personal work of the candidates. There could be no such confidence about workshop-based assignments, laboratory reports and group projects. Assessed group work posed a particular problem; collaboration and team-work were good things to be encouraged, but some students benefited inappropriately by presenting as their own the work of more able or more industrious colleagues.

There had always been a challenge with essays in detecting writing that students did not acknowledge as coming from other sources. This challenge was now much more troublesome in word-processed products in which other people's work could be copied and pasted. Two interviewees spoke of the need for vigilance in spotting work which had been purchased or simply down-loaded from the Internet. There was a need to set imaginative essay and assignment topics for which no standard answers were likely to be already available.

The need for awareness of expanding opportunities for plagiarism and for creative thinking 'to keep one step ahead' of students added to the burden of assessment. One person gave an example where worries about the ownership of reports from a group project led to a recommendation that oral assessment of individuals would also be necessary; realisation of just how time-intensive this would become had led to the proposal being abandoned. The fear of increased plagiarism had acted as a barrier to innovation in assessment.

Two members spoke of their awareness of a Senate Working Party on plagiarism and hoped that The University would provide a clear policy statement and helpful practical guidance.

(c) Objective Testing

The main message here was that 'this chemistry department does not like multiple-choice questions'. One reason for this was made particularly clear by an interviewee.

I am not a fan of multiple-choice questions. In chemistry you have to be able to write formulae and equations, not just recognise the correct ones from a list of possibles.

You have to be able to draw a structure for yourself, not just pick out the right ones from a selection given to you. [2]

This is a strongly stated preference for constructed-response items as opposed to selected-response items. Scrutiny of examples of the short tests (Level 1) showed this clearly translated into actual practice; questions require students, for example, to write the formula for copper (II) sulfate and not to 'identify which of the following is the formula for....'

Another person did feel that multiple-choice tests might well be used at Level 1 where simple recall of knowledge had often to be tested, but this form of assessment would never be appropriate at Honours Level where students had to be asked to understand and to apply. Although it might be possible in theory to construct multiple-choice items which tested more than simple recall, in practice the difficulty of constructing good items to test higher abilities meant that the time to provide 'half-decent' questions was just not worth investing.

Sharp-eyed readers will have noted that the heading of this section is objective testing but that so far only multiple-choice testing has been mentioned. This hints at a fascinating story that the writer would like to tell at greater length than is possible.

In an interchange with one interviewee about the possible influence of computer-based technology in chemistry assessment, the interviewer used the term 'objective testing'. The staff member replied that this term was unfamiliar: 'Does it mean the same as multiple-choice tests?' Later, in one of the extra interviews, the researcher asked why members of the chemistry department tended to treat objective and multiple-choice tests as synonymous. The reply was that this could only be understood in terms of the reactions of chemistry staff (a) to research carried out in the Science Education Unit and (b) to the views of the director of that unit. The director had begun, it was said, by being enthusiastic about the extent to which objective testing should be employed within chemistry; several research studies had then been carried out into multiple-choice testing and these had shown the powerful influence on student performance of minor alterations in the format of the questions, in the ordering of alternatives and in the choice of words. The director had then modified his position on multiple-choice testing; it had been shown to have serious limitations, but other forms of objective testing were still thought to have great potential value. (These included multiple-matching and structural grids.) Members of staff had been looking for simple guidance on objective testing; the kind of guidance they wished was not easy to give in the light of the accumulating research evidence and thus people had tended to select from the Director's message those bits that fitted comfortably within their existing preferences.

Whether or not this account is generally accepted, it appears to this writer to have two implications. (a) It would be helpful if members of the chemistry department were to update their understanding of the variety of types of objective testing, and the strengths and limitations of each. (b) In the light of staff predictions about the likely future impact of assessment technology, it would be useful for them to address what such technology can and cannot do well, and indeed to monitor how these things change as technology inexorably advances.

(d) Examination Stress

It may not be very accurate to describe this as a widespread concern. It is included here because two people clearly showed their concern for students by talking at some length about the effects of examinations on them. The common concern was the recognition that terminal big-bang examinations can be a source of considerable stress. The concern may have been shared, but the conclusions were diametrically opposed.

The first person started from the statement that examinations were very stressful for some candidates. Of course, it could be argued that life itself was stressful and students should be expected to find ways of coping which they would later have to possess anyway. But examinations were a 'very artificial form' of stress and unlike anything ever encountered in 'real' life. It was all very well for some to suggest there was nothing inherently stressful in examinations and that the stress was essentially self-inflicted. This was a spurious argument; stress was none the less real for being self-applied. Some students just did not get the chance to show their true ability because of exam terror. The only answer, in the interests of equity and justice, was to give much greater weight to continuous assessment and to oral rather than written assessment.

The second person began by saying that some staff favoured more continuous assessment out of a misplaced wish to be tender-minded. Of course some students found exams stressful, but what were they going to do when they got out into the 'real' world? There was a distressing tendency amongst today's students when faced with a difficult situation just to say 'I can't do it - it's doing my head in' and then run away from whatever it was. How on earth would they cope with any job with demanding responsibilities? 'The whole idea of keeping your nose to the grindstone is dying out'. Learning to cope with the stress of important examinations was a useful part of professional preparation.

These two accounts should be related to the research evidence. This indicates that although half of all students express negative feelings about examinations (ranging from mild anxiety to suicidal tendencies), one third express similar feelings about continuous assessment ('I feel I can never get out from under it...') Replacing one form of assessment with another will not result in universal contentment. Different groups of students respond differently to different assessment strategies and may be differentially disadvantaged by them. One argument used for diversity in assessment is that disadvantage is reduced by being more widely distributed.

G. GENERIC ISSUES

In this penultimate section, five generic issues are addressed. They differ from the concerns identified in the previous section in that the researcher had them in mind before any interviews were carried out. They are issues which it is predicted will be of relevance in all the four subject areas to be studied. There were again no direct questions on these matters in the original schedules. When, during interview people raised these matters, questions of clarification and amplification were asked. In what follows the issue is first defined and then a summary of the interview data provided. An apology is offered for the attack of compulsive alliteratitus ('The CAs') in issues two to five.

(a) Assessment: The Big Purposes

In the literature, assessment is sometimes described as having three broad purposes: certification, accountability and learning enhancement. At other times, the first two are conflated and two broad purposes are given: certification/accountability (the control function) and learning enhancement (the growth function).

Only two people talked explicitly about the accountability purpose: assessment told us if our courses as a whole were doing what they ought to be doing; assessment procedures provided evidence of whether we were doing our job as teachers effectively.

Two people talked of assessment predominantly in terms of the certification function.

Assessment is a kind of verification that your students have the abilities that you say they have. When you give someone a 2.1 it has to mean something to the public at large and mean something distinct from a 2.2. [4]

Assessment is the central raison d'être for universities and it always has been. We produce graduates and these have to be certificated; we give them labels of approval that people can trust. [7]

Five of the eight interviewees acknowledged this certification purpose, but chose to give greater emphasis to learning enhancement.

My primary aim is education, not assessment. Assessment matters most when it helps students to learn more, to be better educated. [6]

Assessment is about more and better learning. That's why the four short tests get followed up with extra tutorial help. And even summative assessment has a strong formative impact... [2]

(b) Constructive Alignment

Constructive alignment is a concept which has appeared only recently in assessment writing. It refers to that situation in which the aims and objectives of a course (perhaps described as statements of learning outcomes), the methods adopted for teaching the course and the ways in which learning is assessed are all planned together and related to each other so that each guides and reinforces the others. Outcomes, methods and assessment co-exist in a positive, productive relationship i.e. they are constructively aligned.

Only one interviewee spoke of assessment unambiguously in these terms.

I think I may differ from some of my colleagues in that I actually like the idea of structured courses where you have statements of objectives and outcomes and you teach and assess with these firmly in mind. [6]

Another articulated a position which might be described as content alignment or what has sometimes been called curriculum fidelity.

The assessment and the examinations closely represent the course that we have taught; the balance of content in the papers resembles the balance of content in the course we teach. [2]

In what was said by four of the remainder there can be seen a rather different form of alignment, what we might call epistemological alignment. The justification for assessing students by particular means resided in a positive judgement that these forms of assessment were in line with the perceived nature of the subject itself. This was particularly obvious for two assessment methods: lab-work is assessed because chemistry is a practical subject; there are problem-solving exercises and exam questions because chemistry is a problem-solving subject. This alignment-with-chemistry argument is plausible in relation to some changes in assessment practice; it is of dubious value as a defence of degree examinations which are traditional across all subjects. It is now a common criticism of such examinations that they are not constructively aligned with the intended learning outcomes.

(c) Communicating about Assessment

There is a powerful case made in some assessment literature that assessment is only reliable and consistent when the assessors involved are members of a regularly inter-communicating group. One example will help to illustrate the point. There is a naïve belief that reliability in assessment can be achieved through the clear explication of assessment criteria. People are then disappointed when a conscientious exercise in clarifying and agreeing criteria does not remove inter-assessor inconsistencies of judgement. Clear assessment criteria are a necessary but not sufficient condition for reliability; what is needed in addition is that assessors have a shared understanding of what these criteria actually mean. Such shared understandings commonly come about by assessors discussing and arguing amongst themselves about how the criteria are to be applied. (The creation of mark schemes and model answers also helps to attach meaning to criteria.)

Interviewees mentioned an impressive amount of collaboration in the construction of assessment instruments, particularly exam questions. This process was normally based on the circulation of draft questions for written comment; examiners did not get round a table and discuss assessment in general or exam papers in particular. They did of course congregate for meetings of Boards of Examiners, but discussion at these was about the aggregation of marks and decisions about border-line candidates.

There was evidence of shared understanding about standards amongst those members of staff with considerable experience within the department: 'We've been at this for a long time and we know pretty well what people are looking for...'. The importance of effective communication amongst assessors was also underlined by a remark about an external examiner ('It took us some time to train him in our ways of thinking...') and from one recently appointed lecturer ('The old hands seem to know what they are looking for, but I'm still uncertain about standards.')

Some people talked of discussion about assessment within the Teaching Committee; others mentioned conversations and arguments in corridors and in the coffee room. The two groups did, however, seem to be somewhat distinct and there must be some doubt as to whether each group was aware of what assessment issues mattered to the other.

(d) Criterion-referenced Assessment

The distinction between norm-referenced assessment (spreading candidates out to make judgements about their relative standing) and criterion-referenced assessment (judging candidates' absolute standing against predetermined criteria) is a familiar one of long-standing. It is important to ask what these things are. They probably represent polar extreme schools of thinking about assessment or even ideologies to which people claim allegiance in an uncritical way. It is doubtful if they can ever be used accurately to describe any actual assessment scheme or strategy. In practice, pure forms of norm-referenced and criterion-referenced assessment do not exist.

Perhaps the interesting question to ask any group of assessors is the extent to which the more traditional forms of norm-referenced thinking have been influenced by criterion-referenced thinking. The answer for the present group of interviewees is that a minority, perhaps fearing the complexities and uncertainties of hybrid thinking about assessment, appeared to be clinging to earlier certainties; the majority were grappling with the difficulties and confusions of changing ways of thinking.

As an example of the former.....

My philosophy is that honours assessment in particular has to be about grading students; it's about spreading them out so that they can be sensibly allocated to the appropriate honours classification a 1 or a 2(i) or whatever..... Assessment is not an exact science, but it's not usually that difficult to come to a decision that X is better than Y.

[4]

The dilemmas of the latter position can be seen in differing reactions to two apparently similar outcomes. Readers are reminded that a useful diagnosis of norm versus criterion thinking is to ask people what kind of marks distribution they would be pleased with. If they want a nice normal distribution, preferably around a mean of 50%, then they are thinking in a norm-referenced way; if they want a large majority of scores bunched around the upper end of the distribution, then their thinking is criterion-referenced.

Almost all the students were getting nine or ten out of ten[for practical work]. We weren't happy with that....

[6]

Most students get high marks in the short tests, like nine out of ten; this pleases us and motivates them.

[2]

(e) Constrained Autonomy

The idea of an individual academic having complete freedom to do what s/he thought best in assessing students is, and always has been, absurd; individual autonomy is always constrained. The idea of a group of about forty academic staff in chemistry having autonomy with respect to how 'their' students are assessed is not self-evidently ridiculous. However, even here there are clear constraints. Three examples make this obvious. If they wish students to be acceptable to a highly-regarded professional body, then the policies of that body constrain them. If students are eligible to pursue a joint honours degree in, say, chemistry and physics, then the science faculty must have some policy which ensures that two assessment systems are compatible. Students will be awarded a degree not by the chemistry department, but by The University; The University must then exercise its legitimate authority at least over how assessment results are recorded, aggregated and conveyed to a central registry.

All the above constraints were recognised and accepted as proper and reasonable. However, people did have three clear worries about how their autonomy might be increasingly constrained in future.

(i) Constraints might extend into new areas of activity.

There was an acceptance, at times somewhat grudging, that University policy legitimately directed activities in what might be called the administrative and bureaucratic aspects of assessment. Those interviewed would, however, react negatively were The University to tell them what chemistry they should teach, how they should teach it or how they should assess it.

(ii) Constraints might increasingly come from supra-institutional sources.

There was a decisive reaction against the influence of national policy-generating bodies, vaguely 'the government' and explicitly 'the QAA'. Opposition to national bodies was strong when they were perceived as knowing nothing about chemistry; what would be unwarranted interference from QAA might well be regarded as legitimate intervention from the RSC. Similarly one interviewee would be opposed to influence from the Institute of Learning and Teaching, but might react positively to recommendations about assessment from the appropriate subject centre of the Learning and Teaching Support Network.

(iii) Constraints might be imposed in a more directive way.

Attention has already been drawn to the department's pride in its ability to resist some of the pressures associated with modularisation. Here was an externally derived policy which allowed some degree of resistance, redefinition and reconstruction. Future constraints were, it was feared, less likely 'to give us room for manoeuvre'.

H CONCLUDING REMARKS

In the opening paragraph, three research questions were identified. Why do you assess students in the ways you do? When assessment practice changes, why does it change? What issues in assessment emerge as important? Summary answers to these questions will now be attempted; it has to be accepted that brevity produces distortion.

- Chemistry students are assessed in the ways described in Section B and on the whole these ways are not very different from those used for a long time. There have been no compelling reasons for radical change; the assessment systems work well.
- There have been innovations in assessment in existing courses. These have come about from a general feeling that a change in assessment would motivate students to work harder and to learn more.
- New courses have been mounted; changes in assessment were introduced as part of the general process of course planning.
- Chemistry is a distinctive subject and assessment change has been in alignment with the perceived nature of the subject.
- So far external policy requirements have not dictated major changes in assessment practice; this is expected to be different in future. Prescriptions for change will not be welcomed unless mediated through, and endorsed by, a body recognised as chemistry-based.
- Chemistry staff are highly conscientious in their approach to assessment and do not complain about the considerable workload involved. If future developments are to maintain the quality of current assessment practice then two things are needed: (a) increased awareness and understanding of the potential usefulness of computer-based objective testing and (b) more discussion and communication amongst staff about both the reliability of assessments and also their contribution to student learning. Both of these will need staff time.

ASSESSMENT IN HIGHER EDUCATION

SUBJECT AREA II

Philosophy: a descriptive report

Version 2

A. SETTING THE SCENE**(a) The Research Purposes**

The research aims to illuminate the underlying rationale for current practice in assessing student learning; it aims for description and explanation only. The researcher has no role as an evaluator, although the evaluations of others are described. This phase of the research is guided by three general research questions. Why do we assess students in the ways we do? When assessment practice changes, why does it change? What issues in assessment emerge as important?

(b) The Courses

The focus was on assessment at Level 1 and at Honours Level. At Level 1, the department offers six modules annually. These modules are half-year courses i.e. within one academic session a student could take one module from October to January and then a second from January to May.

At Honours Level the department offers five core courses annually; four of these are full-courses and one is a half-course. Honours option courses are either presented annually or in alternate years; in total five full-course options and ten half-course options are offered within any two year cycle. Students may pursue Single Honours in Philosophy, Joint Honours in Philosophy and another subject, or Single Honours in another subject with one or two courses in Philosophy. This is a highly complex provision so, for the sake of simplicity, the treatment within this report assumes a student seeking a single honours degree in philosophy who takes four-and-a-half core courses in year 3 (Junior Honours) and who takes two-and-a half option courses and completes a dissertation in year 4 (Senior Honours). This can be seen as the equivalent of doing eight courses, the dissertation having the standing of a single course.

Later in this report philosophy is described as a high-volume operation. In session 2000 – 2001 the total number of student enrolments (not the number of students as some students take two modules) on Level 1 modules is 1,395. The total number of students pursuing Single Honours in Philosophy is 61 (31 in Junior Honours and 30 in Senior Honours); the number pursuing Joint Honours in Philosophy and another subject is 44 (23 in Junior Honours and 21 in Senior Honours).

(c) The People Interviewed

General approval for the research was obtained from the head of department. She was asked to suggest the names of about eight members of academic staff whom it was recommended should be approached for interview. The request was for the names of people who were heavily involved in some way with the assessment of philosophy students at either Level 1 or Honours Level or both. It was requested that the people should include the following: the

chair of the teaching committee; the examinations officer; the class head for Level 1 philosophy; the class head(s) for Honours Level philosophy; an experienced lecturer with no current responsibility for the design or management of assessment; a lecturer recently appointed to this University, but with experience in assessment elsewhere; a recently appointed lecturer for whom the current post was their first. Nine names were provided.

These nine were approached by letter at the end of October; the letter covered the purposes of the research, the main topic areas for interview and an indication of the time that might be required; enclosed was a protocol making explicit the 'rules of engagement' and covering aspects of approval and anonymity. The letter was followed up by e-mail. Eight people agreed and a date for interview was fixed; one person was unwilling to be interviewed. The eight people were then interviewed in the last two weeks of November; these eight covered the functions outlined above but did not include anyone very recently appointed to The University.

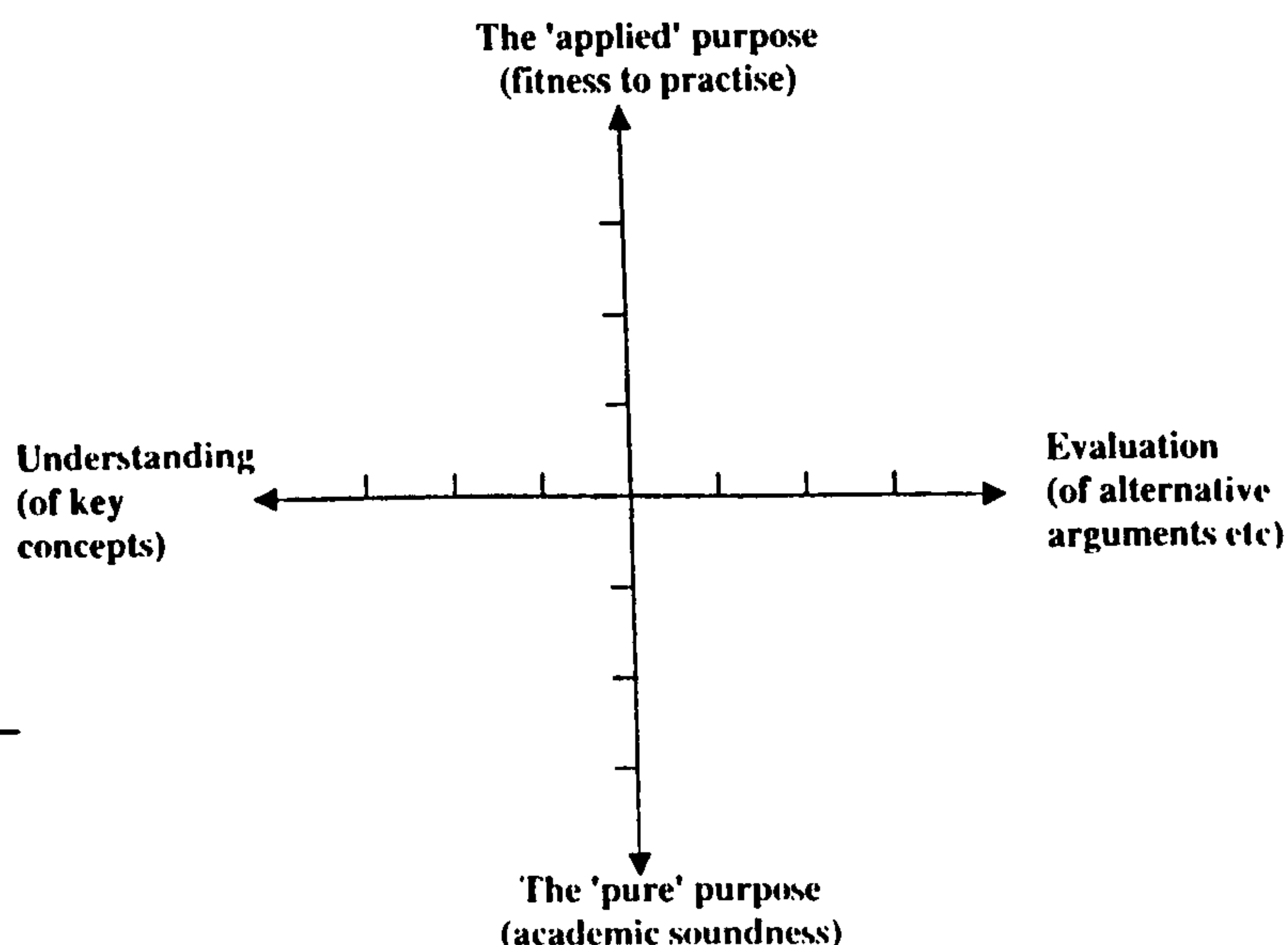
There are currently twenty members of academic staff in the philosophy department; it is **not** claimed that the eight interviewed are a representative sample of this population. Rather they form a purposive sample, chosen because they satisfy the criterion of being well-qualified, through a variety of types of experience, to speak on the topics in the interview. There is no intention to make any empirical generalisations beyond this group.

With each person's permission, all interviews were recorded on audio-tape. Interviews lasted on average for 52 minutes (excluding points in preamble and farewells) and ranged from 44 to 65 minutes. Everyone interviewed had at least twenty years of experience in assessing philosophy students and everyone elected to talk about both Level 1 and Honours assessment. The same interview schedule was used as in the previous subject area interviews.

(d) Philosophy as a Subject Area

In choosing philosophy as one of the subject area to be explored, there is an assumption that it makes sense to treat philosophy as one subject area, i.e. as one discipline albeit containing specialisms within it. This assumption was confirmed in interviews as justifiable.

(e) 'The Four Quadrants' Where is philosophy?



Interviewees were given the diagram above; it was introduced like this. 'It is said that subjects are differently located on this diagram. The axes represent what is emphasised in assessment at honours level. Please do two things, (a) comment on the framework and (b) whatever your reservations, place philosophy where you think most appropriate.'

The first point to be made is that, however much some disliked this simple diagram, everyone had a great deal to say about it. When presented with a request to assess the strengths and weaknesses of a theoretical framework, these philosophers responded with relish and apparent enjoyment. This is hardly surprising; as one said, 'the nature of our subject requires us to be critical, argumentative and indeed combative'.

Several people talked about what philosophy courses **should** be like, rather than what was actually emphasised in honours assessment. It may well be that there is a genuinely close correspondence between what courses should be like and what actually gets assessed i.e. between the rhetoric of course purposes and the reality of assessment practice. Confirmation of this would require not only analysis of what is sought through assessment instruments, but also scrutiny of what gets rewarded in student responses. This goes well beyond the present study.

Much comment focused on the nature of the two dimensions, in particular whether the poles were genuine opposites.

That horizontal axis You can't evaluate arguments that you don't understand. It is a condition of sound evaluation that you understand the concepts involved and you don't really understand the concepts unless you can deploy them in arguments. In philosophy these things should not be opposing modalities; the skills involved run in tandem..... [1]

There is a difficulty with the vertical dimension. Fitness to practise can't apply to entering any obvious profession; very few students will become professional academic philosophers. But students should be fit to practise analysis and argument in a range of employment fields. And the job of practising philosophy needs academic soundness, so practising philosophy is part of the 'pure' purpose. [7]

Placing philosophy somewhere on the diagram was usually done reluctantly. Apart from the meanings to be attached to the labels on the axes, there were three other causes of difficulty. Firstly, different specialist areas within philosophy should be located differently; there were some areas where the key ideas were so difficult that students were less likely in these areas to go beyond attempting to understand the material to making use of it (i.e. to 'doing philosophy'); therefore in these areas assessment was necessarily 'more to the left than the right'. Secondly assessment could be located differently at different times within the same course; as the course progressed the assessment emphasis 'would shift from left to right'. Thirdly, it could be argued that students of differing abilities should show themselves to be in different locations by a qualitative difference in the nature of their responses to the same questions. That meant that essay and exam questions should be set in such a way that the high-fliers would demonstrate that they were towards the right of the diagram; those who were not excellent, but were competent would be in the middle and the 'low fliers' (showing some grasp of key ideas but not actually showing they could 'do' philosophy) would locate themselves to the left. This was the justification of the common double-barreled form of questions: 'What does Plato mean by X? Assess the soundness of his arguments for X.' It should be noticed that underlying each of these difficulties there is an assumption that being good at doing philosophy implies an assessment emphasis to the right of the diagram.

Two people said that as philosophy must embrace all four ends of the suggested dimensions, the only defensible place to locate it was where the axes intersected. The average position of the other six was in the lower right quadrant, at (+1.5, -1.5); within this group of six, five put philosophy within the quadrant and the sixth put it in the upper right quadrant. One person commented that although s/he was choosing the very bottom right-hand corner (+4, -4), really philosophy should be right off the end of the y-axis. This was to correct the tendency of The University to push all subjects up to the top because of a concern for vocational preparation and employability. There was a very big difference between philosophy and, say, management studies and all attempts to make the former more like the latter should be resisted.

Notwithstanding all the reservations expressed, it can be claimed that honours philosophy assessment was thought to put a significant emphasis on the evaluation of arguments rather than the understanding of some defined bundle of key concepts and on a 'pure' rather than an 'applied' purpose. Philosophy was thus perceived as belonging in the 'evaluative and pure' quadrant, but in a less clear-cut way than the researcher might have predicted.

B. STUDENT ASSESSMENT: THE PRESENT

(a) How are students on these courses currently assessed?

All Level 1 modules are assessed in a similar way. Students submit a Class Essay roughly half-way through the module and sit a Final Examination at the end of the module. The Final Examination requires the production of three essay-type answers. The essay and the examination each contribute to the final result in the ratio 40:60. The essay, and each examination answer, is awarded an alphabetic grade (A+ to G); each grade has a numerical equivalent on a 20-point scale; aggregation is carried out by calculating a weighted average of these numerical equivalents and then translating this back into an alphabetic grade (Final Band A to G). A published table relates grades, numerical equivalents and final bands to verbal descriptors: for example, A is 'excellent' and D is 'satisfactory'. Feedback is provided to students from the essay (i.e. it serves both formative and summative functions) but not from the examination.

In describing Honours Level assessment, the simplification described on page 1 will now be employed. The final honours classification (First Class, Upper Second etc.) depends on ten grades, all of which count equally; these grades come from seven courses, one dissertation and two coursework grades. In each course there is a Formal Essay and a Degree Examination which requires the provision of three essay-type answers; the formal essay and each examination answer (all given a percentage mark) count equally towards the overall grade for the course. Students work individually on the dissertation on a topic of their own choosing, with guidance from a supervisor; this results in a written document of between 5,000 and 8,000 words. One of the two coursework grades is the average of marks gained for answers provided in Junior Honours (i.e. year 3) class examinations. The second coursework grade is the average of grades awarded in two Junior Honours projects. In these projects, students work in small groups on a topic chosen from a menu provided and with tutorial guidance; they then provide an oral presentation and a written report. Each student is given an individual mark based on a presentation element (one-third) and a content element (two-thirds). The group as a whole also earns a group mark; this has no summative impact on individuals but may earn the group a fairly substantial prize. A table relating percentage scores, band grades and degree class is openly published. The procedure for determining the class of degree to be awarded is also published and conforms to Arts Faculty regulations: 'this formula is designed to take account of both the average grade and the spread of grades'.

Even with the simplification, this may appear a very complex assessment system; partly this is due to the attempt to describe it within one paragraph. It is described at much greater length in the relevant course documentation. It may be helpful to present the above in tabular form to show the contribution to the final decision of different types of assessment at different times: in the two-year honours course.

	Contribution to final award
From year 3 class examinations	10%
From year 3 projects	
presentation	3.3%
content	6.7%
	10%
Dissertation	10%
From seven courses – one essay in each	17.5%
Seven degree examinations	52.5%
	70%

Two observations seem necessary at this point. Firstly, although the move to giving summative weighting to transferable skills may have seemed radical at the time, the actual weighting given to them is small; oral communication skills account for only 3.3% of the total. (Skills of co-operation and team working do not count formally, but are encouraged by the availability of a prize.) Skills in written communication are also of course transferable, but there seems little chance of their being undervalued in any assessment scheme yet encountered. The evidence is clear that students take seriously those things which are assessed; it may well be, however, that the degree of seriousness is not directly proportional to the assessment weighting. In other words although 3.3% seems minimal it may be sufficient. Secondly, although a significant weight is attached to continuous summative assessment (see C(a)), the weight attached to terminal degree examinations is still more than half (52.2%) and the weight attached to performance under examination conditions approaches two-thirds (62.5%).

(b) What is seen as good?

This section aims to capture what interviewees said in response to the question ‘What in your view is good about how students are currently assessed?’ The intention is to summarise the judgements people made of their own practice, or rather of departmental practice in which they played an active role. The researcher is not writing as an evaluator, but as a recorder of other people’s evaluations.

The question was posed in this way because the researcher was confident that there would be features of the assessment procedures which would readily be identified as pleasing. However, when the question was first asked, of Level 1 assessment, it was obvious from the long pauses before replying, and from what was then said, that the majority were a little uneasy: ‘That’s a question I have never asked myself’; ‘I’ve never thought of it that way before’. Two people began their answers with negative aspects: ‘I wonder if we do anything particularly well, for example I don’t like....’ and ‘We only use two assessment methods and I’m not entirely happy with either....’ However they started off, each member of staff then went on to provide a considered answer to the question ‘What are the strengths and limitations of current assessment practice?’ Perhaps this is the form of question that

philosophers characteristically like to pose and answer. When asked what was good about Honours assessment, respondents found it much easier to identify and describe positive features. To conclude that there is more deserving of praise in honours assessment might be unwise; a mere sequence effect in the interviewing process may be at work.

Level 1

Three things were said, each by a majority of staff, to be good: the essay, the degree examination and the relative weighting given to each. Two aspects of the essay received favourable comment. Firstly, it provided a good opportunity to provide constructive feedback to students early on in the module i.e. it played an important role in formative assessment. It was repeatedly said that many students found difficulty in grasping the nature of philosophy and in presenting their thinking in the ways expected within a philosophical essay. All essay markers were accustomed to writing extensive comments and students were encouraged through these to seek further help from the tutor/assessor during his or her 'office hour'. The main thrust of the feedback varied of course with the needs of the students, however there did seem to be interesting differences in what individual assessors would emphasise in their feedback.

Students have to be put on the right track. They think we want airy-fairy waffly stuff. We want dead simple, crystal-clear stuff, succinct and concise. They can learn this from the feedback we give on essays. If they were only to learn it later the hard way, through the exam, it would carry too great a penalty. [2]

I like it when we are free to look for originality, without worrying too much about whether they have got hold of the received wisdom and perhaps got all tangled up in it. I can say to them 'you've got it all wrong, but in an interesting way'. And then I pause and think 'But have they indeed got it wrong or just approached it from an unexpected perspective?' This sort of thing can happen with the essay, less so with an exam. [8]

Secondly, it was seen as good that the essay also had a summative role. There were three reasons for this. It motivated students to take essay-writing seriously; it was more authentic than producing essay-type answers under time-limited exam conditions as it allowed students 'to develop an extended coherent line of argument'; some students who under-performed in formal examinations could be 'pulled through' by their better performance on the essay

The existence of a degree examination was seen as 'a good thing'. The end of module examination was more rigorous than other forms of assessment; it required the bringing together of knowledge and understanding from different parts of the course, it was less dependent on secondary sources at the time of writing; it offered less scope for plagiarism .

At least with an examination you can be sure it is their own work: the student has to turn up, sit down and do it, without collusion and without props. [4]

The 40:60 ratio for weighting the essay and the exam was seen by the majority as appropriate. It was preferable to the 50:50 split that had earlier been the case; this gave too much importance to the early essay. Two of the eight would have liked the weighting of the essay reduced (perhaps to give a 25:75 ratio); it was just too difficult for students to produce a good essay as early as six weeks into the module, particularly when this was in the first philosophy module to be tackled.

Honours Level

There were two general aspects of Honours assessment each praised by the majority of members of staff: the diversity and balance within the overall assessment scheme and the 'quality' of the final examinations.

Diversity in the modes of assessment deployed was judged to be pleasing: each method tapped rather different but relevant skills and abilities. When any course had diverse aims it was unlikely that any one assessment method could validly assess all of them. Different groups of students within the Honours population were advantaged and disadvantaged by different assessment methods; thus making use of a range of methods was seen as fairer overall. Not only were the assessment methods appropriately diverse; they had relative weightings attached to them that were seen as highly satisfactory; no criticism emerged of any aspect of the weightings.

Given this diversity, it pleased interviewees that final examinations retained an important place and that these examinations were of high quality. Examinations possessed important positive features, some of these as already indicated above: they required students to integrate learning from different parts of the course; they could elicit performance at a higher level than previously; they gave less scope for student dishonesty than some other methods. However, staff were not just re-stating the benefits of examinations as a method, they were singling out for praise the ways in which, generally and for the most part, their colleagues conducted all the procedures relating to examining: exam questions were carefully constructed and scrutinised and were thought to be 'good' questions, demanding but fair; assessment criteria were explicit; because a team of markers were involved, results were discussed and different opinions accommodated; marking and review procedures ensured an acceptable degree of inter-assessor reliability.

Other individual features of the assessment programme were singled out as pleasing. Each of the following was identified by three members of staff: the weight attached to essays produced in non-examination conditions; the dissertation; the project presentations and the 'carry-over' from the Junior Honours year.

Formal essays carried one-quarter of the weight in each of the seven full courses required for single honours in philosophy. Giving this summative weighting to formal essays was seen as having greatly improved the seriousness with which essay writing was tackled and even the punctuality with which the products were submitted. Essays helped those candidates who found examinations particularly stressful.

The dissertation was praised. The topic 'belonged' to the student, although guidance in its choice was available; whereas in essays there was a lot of duplication, because students co-operated in their production, the dissertation had to be very substantially the work of one student alone, working independently. Interviewees said that although 'the dissertations were on occasion quite remarkably good' they did not wish any greater weight to be attached to them.

The project and its presentation tapped into a different range of skills. They did require co-operation, collaboration and effective team-work and significantly they required skills in oral delivery and a greater competence 'in thinking on their feet'. Those who praised the project-presentation saw it as being clearly in line with a valuable aspect of philosophy: 'students' ability to talk philosophy, rather than write it, is very important indeed'.

The carrying-over of the two course-work grades from Junior to Senior Honours was thought to encourage serious application to study earlier in the course. Again, however, there was

clear agreement that this should not be given greater influence than it currently had (i.e. two grades in ten).

There were two further features of Honours assessment each praised by a single person. Firstly, the project presentations introduced an element of student comment which helped inform staff judgement i.e. there was some peer assessment, but without any summative role. Secondly, working with relatively small tutor groups provided valuable opportunities for that informal and diagnostic assessment which is integral to effective teaching.

Everyone was well aware of the limitations of each assessment method: examinations did not produce good work from those who found them seriously stressful; essays could be depressingly similar due either to co-operation in their preparation or to routine regurgitation of received wisdom; presentations ranged from the tedious reading of an essay to an engaging dramatic performance of trivia ('the effective communication of what is not worth communicating'); the topics for dissertations most appealing to students (for example in applied ethics) were often the most intractable.

There were two criticisms of Honours assessment which were not related to the limitations of particular methods. The first was that the system as a whole tended to produce work from students which was respectable but dull; this point is developed further in Section F(a). The second was that Honours assessment effectively and reliably identified the high-fliers and the earth-bound i.e. it worked at the top and the bottom of the student ability range, but it failed effectively to differentiate amongst the largest group – those somewhere in the middle.

C. STUDENT ASSESSMENT: LOOKING BACK

(a) Introduction

Interviewees were asked to identify what they saw as significant changes in how students have been assessed. It was decided beforehand not to restrict people to some arbitrary period of time, but to encourage them to reflect on the whole period over which their experience extended. Interviewees were then asked why they thought a change had occurred. It was anticipated that this would be a difficult question to answer and so it proved. Change usually has many subtle and interacting causes; it is possible to be involved in change and not to know why it is happening; it is possible that one knew why at the time but cannot now recall the reasons. However, no apology is necessary for asking the question; people's perceptions of reasons are arguably as important as the 'real' reasons, even if it could be agreed that these are ever accurately knowable.

Some clarification of terms as used by the interviewees and as used in this report is necessary here. Firstly, some staff referred to continuous assessment and some to continual assessment. Although continual (in the sense of frequently recurring) is perhaps more accurate than continuous (connected throughout in time), conventional current usage favours continuous. In what follows the term continuous assessment is used. Secondly continuous is taken by the writer to be the opposite of terminal: continuous assessment takes place during a course; terminal assessment occurs at the end of the course. These terms have to be very clearly distinguished from formative and summative. Formative assessment is taken to be assessment which provides the student with some kind of feedback intended to influence their future performance; summative assessment is assessment which provides grades or marks which are taken into account in making end-of-course judgements of success and failure (or of levels of success). Formative assessment is usually continuous, but continuous assessment may be summative as well as formative. Terminal assessment can have no formative impact during

the course, but could well have a formative influence on subsequent, after-the-course learning.

(b) Significant changes at Level 1

There was general agreement that assessment methods had not changed very much at Level 1; there had been more change at Honours. Assessment of student learning was carried out through essays and essay-type questions within formal examinations. These were seen as appropriate methods 'given the nature of philosophy'.

Change has not been drastic. Assessment has always been by essays and exams in my time here.... and that's not wrong given what philosophy is all about. [2]

Although the actual methods used have not changed much, there were still said to have been significant changes in assessment. The one described by a majority of staff was the increased amount of continuous summative assessment. In the past, philosophy at Level 1 had been a year-long course in which there had been two essays and two class examinations; success in these conferred eligibility on students to sit the final degree examination, but grades on the four components had no summative role. After the advent of modularisation, a student doing philosophy at Level 1 would normally take two consecutive modules, each being a half-year course. Within each of these, as described earlier, there came to be one class essay (somewhat shorter than before) and one examination, both of which made a summative contribution to the final grade for the module. This change had resulted in less opportunity to support student learning through constructive feedback.

Twelve weeks is just not long enough for student development; no progressive formative assessment is possible in these short modules. Every exam is now a degree exam. Conversation between the marker and the student has been lost. There was a campaign with the external examiner and the students to go back to year-long courses for this reason. [4]

The modular system, although it may have some advantages, cannot compensate for that weakness, for the loss of formative assessment. [5]

Two other significant changes were identified, each by three people. The first was the loss of exemptions from the final examination; the possibility of gaining an exemption had been an effective motivation for some students to work harder throughout the course. The second was the introduction of anonymous marking, seen as a desirable change.

(c) Significant Changes at Honours Level

There was general agreement that the biggest change had been a move to more, and more diverse, summative assessment. The main features of this were the attaching of summative weighting to essays (written outwith examination conditions) and the introduction of project presentations.

There has been a significant change in the balance of assessments; whereas essays were previously part of the teaching process, they are now important elements in the examining process. [3]

The presentations in year 3 widen the skills we now assess – fair enough; the present summative impact of this is relatively low, but it's OK. I wouldn't want them to count too much. [7]

It was very noticeable that although the increase in continuous summative assessment was just as frequently mentioned as for Level 1, the level of objection to the reduction in opportunities for formative input was much lower. This is presumably due to the length of the course; the Honours course spreads over two years, Level 1 courses are over in half a session. On the whole there was cautious approval for continuous summative assessment at Honours Level; it helped to ensure students worked more consistently throughout the year and reduced a little the stress associated with terminal 'big-bang' examinations.

Changes in three other aspects of assessment were seen as significant: the nature of exam questions and essay topics, procedures for aggregation and anonymous marking.

There had been an important alteration in the nature of the questions students were required to tackle. Those who saw the change as unfortunate thought that questions now elicited less creative thinking from students; those who saw the changes as a good thing felt that questions were now more 'professional' than they had been. This issue is discussed in Section F (a), but two quotations illustrate the different perspectives.

Questions are becoming much more stereotyped; they don't give students room for originality. There are doctrines on things and well-known refutations of these; our questions now expect students to reproduce these. In the old days we asked questions which were often whimsical and oblique. 'Does justice pay?' rather than 'Discuss Smith's doctrine on whether justice pays'. [8]

The character of the exam papers is now much less eccentric than it was. I know the same thing can be seen as either good or bad. I tend to see a diminution in eccentricity as an increase in professionalism – in both our teaching and our assessments. The overall effects are positive rather than negative. [3]

Changes in aggregation procedures, i.e. in ways of deciding the final honours classification, were seen as significant. The faculty scheme for recording the results of assessment resulted in an array of marks from which there was an algorithmic mechanism (capable of being done by a computer) for generating the degree class a student should be awarded. Only in cases which were both borderline and unusual could members of a board of examiners exercise any discretion. Although the main reasons for the introduction of such a procedure were appreciated (to simplify decisions for students pursuing joint honours degrees and to make decisions more easily defended in cases of appeal) there were misgivings about what was seen as the decreased importance attached to academic professional judgement. Any mechanical scheme for aggregation made assumptions about the accuracy of assessment of the components and about their commensurability that were intellectually suspect.

The change to anonymous marking at honours level was mentioned by two people as a significant improvement.

Anonymous marking was considered by the department to be a good step forward at Honours Level. When you know the student it is very difficult to avoid the halo effect. And anonymous marking is actually less stressful for the marker – you feel you are assessing the product rather than the person. [6]

(d) Reasons for these changes

At the outset there was a hope that people would identify the change they saw as most significant and then identify the reasons they saw as operating in that single case. Either this hope was naïve and misconceived (given the nature of causality in human affairs) or the interviewing technique was inadequate to the task. What happened, particularly at Honours Level, was that two or three significant changes were described and then two or three influencing factors were identified. It was not possible to estimate the potency of any one factor in relation to one specific change.

The significant change in assessment practice at Level 1 was ascribed to a relatively straightforward cause: the faculty policy which embedded the institutional pressure for modularisation. Change in assessment had been required because of a required change in course structure. The assessment change which had been made to meet this required change in course structure was an alteration in the timing and the status of the essay and the final examination. There was no change in the actual methods of assessment; these methods had served the department well in the past, they were perceived to be appropriate given the nature of the subject, so they continued to be deployed.

The picture at Honours Level was rather more complex. Why had there been a move to more, and more diverse, continuous summative assessment? In the run up to the Teaching Quality Assessment exercise, it was recognised that one criterion that would be used in this was 'variety in the forms of assessment used'; this pressure for diversification had not been resisted by the department, indeed there was a general feeling that it would contribute to the overall fairness of assessment. (Students who did not shine in one method of assessment might well do better in others.) The TQA was also thought to have influenced departmental thinking on what should and should not have summative weighting; there was a recognition within the department that continuous assessment was fairer for those students unable to do themselves full justice under examination conditions and also that it should help to motivate students to do more work throughout the year. Although continuous summative assessment had respectable educational justification, it was also seen by some as 'just something in the air at the time – a feeling that it was fashionable'.

The precipitating cause of assessment change seems to have been a supra-institutional policy, mediated through the institution (specifically the Quality Assurance office) and crystallised in faculty policy. However, the department had found a way of responding to external pressure which they could readily defend on educational and equity grounds and which they could construe as being in line with the nature of the subject. External policy requirements were not always viewed favourably; whereas policy on anonymous marking was thought to have had positive effects, that on aggregation had not.

At Level 1, extra-departmental policy (on modularisation) had resulted in assessment change that was seen as unfortunate; at Honours Level such external policy had produced a number of assessment changes that were, on balance, seen as beneficial.

D. STUDENT ASSESSMENT: LOOKING FORWARD

(a) Changes wished

Interviewees were asked what changes they personally would like to see in how students were assessed, assuming that the conditions for introducing such changes could be made favourable.

Level 1

The generally expressed view was that the present assessment procedures were 'pretty good given the constraints which apply'. As long as Level 1 philosophy was 'a large volume operation – which it should be' it would be difficult to improve things, given that there was no such thing as a perfect assessment system anyway. Several said that the only thing which would allow any development in assessment would be a reversion to a year-long course because 'short modules militate against both diversity and innovation in assessment'.

Nevertheless, four changes were seen as desirable. These were as follows: more informal and formative assessment (five people), less dependence on essays and examinations alone (three people), encouraging more originality within essays and examinations (two people) and reducing reliance on graduate teaching assistants (one person).

Staff wished to see their students for more time and in smaller groups so that they could carry out that informal assessment which is integral to effective teaching; they also wished students to submit more work which could be formatively assessed. Most wished students to submit additional essays which could then receive helpful feedback about what they were doing well, what the significant weaknesses were and what was necessary for future work to be more highly rewarded; two felt that in addition such essays should be given a mark (to let them know quite unambiguously the standard of their performance) but one which would not count towards the final grade.

Those who wished to reduce dependence on essays and essay-based examinations identified ways of doing this: short diagnostic tests and oral, 'viva-like' examinations. These would have to be carefully designed and implemented.

I'd like to break out of the essay/exam pattern; there are other things we might do; we should experiment with a range of other modes of assessment. For example, there is a place for shorter tests, to test the more basic learning outcomes. The orthodoxy here is that short tests mislead students about the nature of philosophy and distort their views as to what's wanted. This is not necessarily so..... [5]

We need more diagnostic testing; some short tests could be done at an early stage and quite quickly – I can see how to do it for my own courses. But it would need a general departmental effort and careful thought. There would be a danger of giving the students the wrong expectations. [6]

Oral exams are good back-up devices. If they are thorough they are appropriately demanding of students – and exhausting for the examiners. It is important not to be misled by students who are merely good at talking. Oral exams can also show up the inconsistencies between examiners.... [7]

Two people wished to see the development and use of essay-type questions, whether as formal essays or in the examinations, which would encourage more originality and creativity from students. The standard questions tended to encourage reproductive rather than productive learning; students were rewarded for regurgitation rather than for showing their ability to think when confronted with the unexpected. This was a plea for question setters to be more imaginative, but it might require more than this. It might be necessary to free up essay/exam time by assessing 'the basics' elsewhere (perhaps in short tests) so that there was 'elbow-room' to test the higher cognitive abilities of application, synthesis and evaluation within the essays.

Graduate teaching assistants mark essays and may, after a year or two, also mark examination answers; their assessment activities are supervised and monitored. One member felt that the

department's increasing reliance on GTAs should be put into reverse. This was not because of any doubts about GTAs' conscientiousness, which was very high, but because of student perceptions that their marking was less reliable and their grading was less authoritative than that of experienced academic staff.

All the things that staff desired were linked to a clear comment that what was needed to make them feasible was more staff time. Additional time was necessary to give more opportunities for informal assessment, to give more constructive feedback to more student products, to design effective diagnostic tests, to conduct reliable orals, to devise new types of essay questions and to rely less on graduate teaching assistants. There was a belief that diversification in assessment methods elsewhere had often led to the over-assessment of students (because people always found it easier to add new assessments than to excise existing ones) and to an unacceptable increase in the assessment workload burden for staff.

Within a department that had always attached great importance to teaching (under which heading interviewees put all but the administrative aspects of assessment), the ever-growing stress on research activity tended to force the demotion of teaching. Pressure on the time that staff now felt able to devote to teaching was a factor inhibiting change in assessment methods. Although the people interviewed in this study could all identify developments they thought desirable, they regularly expressed doubt as to whether there was the general, departmental will to devote extra time to assessment. This was not to be seen as complacency, but rather as an indicator of stress within the system.

Honours

Staff had less to say about changes they would like to see in assessment at Honours Level. Again the general view was that the present assessment system was sound, given the resource constraints.

The most generally desired change, as at Level 1, was for more assessment that was informal, formative and diagnostic. This, for most, would require the re-instatement of individual one-to-one tutorials, but staff were very clear that they did not wish tutorial or seminar performance to be summatively assessed. More informal assessment should also occur in project and dissertation preparation and via short, diagnostic tests.

Two people wanted more weight attached to the project presentations; students should be given more encouragement 'to talk philosophy'. However, as a precursor to this, it would be necessary to do more to explicate the assessment criteria and to reach agreement as to what these criteria actually mean.

We should make more use of the presentations. This means giving them more weight, but not just that. We have to train students earlier in the year in presenting arguments. And there is rather too much uncertainty amongst academics about what they should be rewarding.... [3]

The pleas for more open-ended and more imaginative questions to encourage original thinking in students was repeated for this stage.

(b) Changes that may be required

Staff were invited to speculate about changes in assessment practice that might be required of them, whether they would welcome these or not. Most people started off by saying 'Nothing very definite' or 'I don't know of anything' – and then went on to talk at some length. All the changes that people foresaw involved extra-departmental pressure and policy; perhaps the

form of the question made it unlikely that people would predict that departmental thinking would require them as individuals to change. It was very noticeable that when speculating about future changes people talked much more of policy than they had when reflecting on past changes. Policy requirements would come either from some national, supra-institutional source (such as QAA) or from some institutional, extra-departmental source. Whether the latter was The University (i.e. central, senior management) or The Faculty (of Arts) was not easy to disentangle. In the hierarchical structure that operates, policy usually comes through the faculty; where it starts from is not then seen as a matter of much importance.

There might be a requirement to attach more weight to the assessment of transferable skills, particularly skills in oral communication. There was disagreement on the desirability of this. On the one hand, students might take more seriously the need to develop skills in expressing arguments. On the other hand there were huge difficulties in making reliable judgements of communicative capability, ones in which different assessors made consistent distinctions between the style and the substance.

There might well be an externally imposed requirement to give even greater weighting to assessments of all types through continuous summative assessment. However, a move to greater diversity in assessment combined with more continuous summative assessment could result in 'a further reduction in the integrity of our assessments'. When fair and reliable assessment becomes increasingly difficult to sustain, there is then a pressure to revert to those assessment methods in which cheating is more difficult i.e. formal examinations. There are two potentially conflicting forces: one seems to require more diversity in assessment and the other militates against it. One could envisage different policy demands, from different sources, requiring departments to try to move in different directions.

Changes to assessment practice might also be required if staff numbers in the department were allowed to fall, i.e. if the student/staff ratio were allowed to get higher. The existing assessment strategy depended for its successful operation on the 'assessment expertise' of the existing pool of staff and on the time available to be devoted to conscientious assessment: 'Assessment is driven by how much time, from suitably qualified staff, can be devoted to it'.

Influences such as those outlined made some people think that it was inevitable that in future The University would come to act, in relation to assessment, in a much more directive way than had been necessary up to now.

Senior management in The University may direct more. There is a general feeling that central management is becoming more powerful and intrusive. They may well infiltrate into the assessment field as well.... [8]

Although some felt that The University might intervene in assessment by prescribing directly how students should be assessed, other felt it more likely that the influence would be an indirect one via, for example, policies on pass-rates and Special Needs cases. Both of these deserve some amplification. Firstly, there was a perception that standards within philosophy were being pushed down because of the wish of central managers to pass more students, 'to allow them through'. There had been complaints in the past, it was alleged, about both pass-rates and completion rates in philosophy; if these were to continue or to intensify, then the standards applied within the assessment process would be forced to change. Secondly, and this view was carefully worded to avoid the risk of its being misunderstood, University policy on Special Needs cases was beginning to have an impact on general assessment practice. Management were very keen to be seen as inclusive; there was increasing input from Special Needs Advisers (and emerging legislation) in support of this policy and in relation to assessment. A candidate with special needs could be supplied with a scribe, a dedicated computer and specially presented examination papers; in some cases an oral examination was added on to the written components. Undoubtedly staff had great compassion for individual

cases and went to great lengths in making special provision, but if the definition of 'special' needs came to be too wide, then a new problem of inequity in assessment would arise. The time that could be devoted to student assessment was limited; the more time that was devoted to cases of special need, then the less time that could be spent on those not so defined. Assessment was required to be reliable, valid and fair - and fairness was becoming increasingly difficult, both to define and display.

Staff in philosophy might be required to seek greater economy in their assessment procedures through the use of computer and information technology. There was room for experimentation here, but it was thought by most that the subject did not lend itself to 'objective' assessment; it was not theoretically impossible, but it was hugely difficult (and very time-consuming) to generate items which effectively assessed the higher intellectual abilities.

It was thought relatively unlikely that any supra-institutional body would prescribe changes in assessment, although they might well affect the general climate in which thinking about assessment took place. An earlier fear that QAA would require changes in assessment practice seemed to have receded. Benchmarking would probably not have any very significant influence: 'the benchmarks are very broad - any respectable course in philosophy already satisfies them'. There was in philosophy no professional body which would press for change in assessment, or which could veto any change suggested by a philosophy department. It could be that the Philosophy Centre of the Learning and Teaching Support Network would make interesting suggestions in future as to how assessment might develop within philosophy.

E. ASSESSMENT: FACTORS INFLUENCING CHANGE

In Section C(d) there was some consideration of the factors that interviewees gave as being important in past assessment changes. In Section D(d) there was an indication of the factors that people thought would become more influential in the future.

After attempting to elicit each participant's own views on factors influencing assessment change, the interviewer presented each person with a list of the views of others and asked for responses to these. People were asked (a) to give their personal view of the importance of these factors in assessment change in philosophy in the past and (b) to predict which would become more influential in future. It was very noticeable that interviewees were unwilling to give rapid responses; often there were long, thoughtful pauses before people committed themselves. The results are summarised below.

Factors influencing past change in assessment

Factors for which the most frequent response was 'very important for us'

Change in course structure

Factors for which the most frequent response was 'of some importance for us'

Concern for transferable skills
National policy required change
General concern for the quality of student learning
External examiner recommended change
Increased number of students
Faculty policy required change
Institutional policy required change

Factors for which the most frequent response was 'not important for us'

One enthusiast persuaded us
Awareness of assessment trends elsewhere
Students pressed for change
Increased diversity within the student population
Availability of assessment technology
External professional body
Impact of educational/assessment theorist

Factors predicted to become more influential in future

Institutional policy
National policy
Faculty policy
Pressure from students
Awareness of assessment trends elsewhere
Availability of assessment technology

The above tables are derived directly from the number of boxes ticked. Interviewees regularly gave comments on their choice of response. The main points they made about each factor are given below; the factors are dealt with in the order in which they appear in the first table above.

1. There was general agreement that the major influence on assessment change had been the required changes in course structure at Level 1. It was inevitable that assessment had been restructured as part of the process of course planning. There had been no corresponding change in course structure at Honours Level.
2. The assessment of presentation skills in Junior Honours projects was introduced out of a concern to pay more attention to transferable skills. There was a difference of opinion about whether this was giving explicit recognition to something which had always been important ('transferable skills have always been there within philosophy') or whether it was taking philosophy off in a new and undesirable direction ('we did it because we thought we had to, rather than genuinely accepting it was our business to engage in this kind of thing').
3. When national policy was said to be an important influence in bringing about change, staff usually had in mind the impact of the Quality Assurance Agency in the run-up to Teaching Quality Assessment (which had resulted in the philosophy department being judged excellent).

TQA required us to think about how to introduce more diversity in assessment (like the projects) and to have more continuous assessment. We felt we were being told to mend our ways, because examinations were no longer fashionable. [5]

4. People seemed not to be saying that a particular change X had been introduced out of a conviction that it would enhance student learning, but rather that a concern for what students learned permeated all discussion of assessment. It was the crucial factor, for example, in producing the negative reaction to the assessment changes required by modularisation; it was also the motivating force behind the wish for more opportunities for informal and formative assessment. People felt that this factor must be seen as important otherwise they might be open to the charge of not caring enough about the link between assessment and student learning.

But look at how you might interpret this. If I said that a concern for the quality of learning wasn't a factor, you might be misled. We have a lot of concern. But if we thought our assessment practice was basically sound, then concern for student learning was not a factor influencing change. Don't interpret a lack of concern as complacency. [1]

5. The influence of external examiners was perceived as important, not just in relation to decisions on individual candidates but more generally on how students were assessed. However, although staff could recall many examples of changes in detail on specific papers and questions made in response to the wishes of external examiners, they could not identify any major change in assessment methods that could be attributed to the influence of an external examiner. All departmentally suggested proposals for change had been thoroughly discussed with them, but the role of external examiners in relation to such change was reactive rather than proactive.

External examiners might, it was said, advise a change in assessment methods, but they could not require it.

External examiners interpret their role differently. If we were to be stuck with a martinet who made some unrealistic demand, then we'd find a way around it. [7]

6. An increase in the number of students was seen as important, even though it had not forced any radical change in assessment practice. Bigger numbers had meant that the required length of essays had been reduced and that there was rather less re-marking and double-marking than previously. Perhaps, it was suggested, people were also giving less feedback on essays.
7. In responding to items about policy, it was frequently mentioned that it was always difficult to say where policy came from. Staff within the department were unsure whether, for example, the faculty had originated a particular policy or acted as the channel for an institutionally derived policy. Similarly, one could not always be clear whether The University was formulating a policy or acting as a channel for national, or even international, policy. For example, modularisation may have started as an international concern, then become an institutional pressure and then finally translated into a faculty requirement.

Faculty was perceived as normally playing a reactive role i.e. in responding to changes suggested by departments. However, in the run-up to Teaching Quality Assessment, it had become more directive.

Interviewees thought it would be somewhat improper for Faculty to require that particular assessment procedures be used; it was seen as reasonable, however, that Faculty had required a change in marking systems for Level 1 (from a percentage to a 20-point scale) in the interests of inter-departmental comparability and intra-faculty uniformity.

8. Again with institutional policy there was a problem for individuals in deciding where policy came from. It was unclear, for example, when the Quality Assurance Office in The University was originating policy for the Senate to endorse and when it was conveying policy from an external source.

The University was said to have generated policies on the use of marking scales, letter grades, grade descriptors, methods of aggregation and anonymity in marking.

Central management was said to have created an institutional climate in which it was considered politically incorrect to say that any student had failed a course at Level 1; instead they get grades E, F and G. (Although 'everyone knows that you have to get a D to pass'.)

Some of those interviewed reflected on whether or not policy required something of them (to do X and not Y) or established a climate which made it more likely that they would themselves choose to do X rather than Y. (It will be necessary to return to this point in a later discussion relating the conceptualisation of policy and the nature of causality in human affairs.)

We were actually required to do none of these things; but when other people change around you in a policy-driven fashion, it can be dog-in-the-manger to resist, even although you regard some of the changes as malign. [1]

It is interesting that several members of staff said that the university policies indicated above had, on the whole, produced a desirable consistency rather than an inappropriate uniformity. They feared that the balance was likely to shift in the future.

9. The influence of an enthusiast for change had been unimportant, except perhaps in relation to the introduction of project presentations.
10. Awareness of assessment trends in philosophy elsewhere came indirectly from external examiners or from the interviewees' experience of external examining. Two people said there had been some additional scouting around but 'we didn't discover anything compellingly better' and 'there are precious few ideas about innovative assessment within the philosophical community'. The Centre for Teaching Philosophy and Religious Studies (part of the Learning and Teaching Support Network and based in Leeds) might well change this.
11. All changes in assessment had been thoroughly discussed with students. Students had not pressed for changes in the methods by which they were assessed. It was a characteristic of students to ask for very different things at different times and some of their suggestions were 'just not practical'. Several people did say that it seemed likely that students would press for assessment change in future.

There are two reasons we may have to respond more to students on assessment. Because of serious impoverishment, they need time to earn money and perhaps we will have to assess them more flexibly. Secondly, they are becoming a more demanding clientele; they may not continue to accept some of the old practices. The pressure from students as customers will mean that all our methods, procedures and formulae are clear-cut, transparent and easy to defend. [3]

This point should be borne in mind in a later discussion of the growing pressure to re-conceptualise and de-couple the linked ideas of reliability and validity and an apparent trend to stress the former at the expense of the latter.

12. Awareness of the increasing diversity within the student population had resulted, not in a change of assessment methods, but in a need for extra support to help some students to be ready for assessment. A few staff thought that increasingly diverse student abilities had contributed to 'dumbing-down' within philosophy assessment.
13. The availability of appropriate technology had affected the recording of assessment results and the statistical analysis of them; it had also made new forms of plagiarism

possible. There had as yet been no moves towards computer-based assessment, with one exception.

I do use Internet exercises for training students in self-assessment. There are various quizzes on the Net to test knowledge of use of language. Computer-based assessment does tend only to be appropriate for the elementary aspects of our subject. What I do is formative practice in self-assessment; I wouldn't want to use it formally, summatively. [7]

14. In philosophy there was no professional body, comparable with say the Royal Society of Chemistry or the General Medical Council, which could influence how students were assessed.
15. There had been no impact on assessment practice from educational/assessment theory or theorists.

Interviewees identified six factors that they predicted would become more influential in the future: institutional, faculty and national policy, pressure from students, assessment trends elsewhere and assessment technology. It is considered significant that external policy, whatever its source, dominates this list.

F. DEPARTMENTAL CONCERNS

In this section four concerns are identified: originality, integrity, standards drift and workload. The researcher had no preconceived idea that these would emerge as important to the people interviewed; questions about them were not included in the original interview schedule. In labelling these 'concerns', the word is used to mean a matter of interest where there is an undertone of worry or anxiety; conventionally in research reports these matters would be called emerging issues i.e. important subjects of debate. The data suggest that these topics have been the subjects of only limited debate so far.

(a) Originality

Implicit in many comments from staff was their recognition that the nature and quality of student learning was in some way related to what was sought and rewarded in assessment procedures. One example of this was originality in student thinking; students displayed less originality in their written work than staff wished to see and this, staff said, must be linked in some way with how that work was being assessed.

There was little or no scope for originality when students were required by the question or topic to display understanding of some standard position, doctrine or line of argument. Scope did exist when the form of the question required students to apply their understanding to some unfamiliar topic or problem. There were two distinctions that it was useful to make. Firstly, original thinking was not to be confused with missing the point entirely or the merely bizarre; creative thinking was not a leap out of ignorance into the dark. Secondly, creative thinking about the substance of philosophy was not the same as imagination in presenting that substance; some students devised highly creative, and enjoyable, ways of presenting their projects and there was a risk when assessing these that pleasure in the style could mislead about the quality of the content.

There was some difference of view about what had been the significant aspect of assessment in reducing originality, more emphasis on continuous summative assessment or a change in

the nature of the questions asked. Continuous summative assessment was thought by some to have made students overall much more cautious about 'going out on a limb and taking risks'; they tended to play it safe and give back to their tutors what they had received from them.

They have precious little opportunity to mess-up, to learn from their mistakes. They no longer have any chance to be adventurous; they don't risk tripping-up when it could affect the rest of their lives. [2]

Others believed that the significant factor was that the topics set and the questions asked were much more predictable and less open-ended than in the past. (See Section C(c)) In the past lecturers had been responsible for the teaching and assessing of their own specialist enthusiasms; they had felt free to set questions which could be seen as distinctive, imaginative and likely to stimulate creativity in students. Such questions could of course also be seen as quirky, whimsical and idiosyncratic.

What some people saw as a regrettable decline in academic freedom, others saw as a desirable increase in professionalism; the increased influence of course teams had curbed individual eccentricity. It does seem highly unfortunate if academic professionalism is such that it can only be gained at the expense of student originality. This suggests that academic professionalism itself is in need of some re-definition.

(b) Integrity

The word 'integrity' was often used in interviews. It was used in two different ways, the first in relation to assessment methods which did not give opportunity for student dishonesty and the second in relation to assessment procedures which reduced the likely impact of assessor inconsistency and bias.

A majority of those interviewed were concerned about plagiarism in some form or another. There had probably always been a minority of students guilty of some dishonesty or cheating during assessment and the present situation should not be depicted as a crisis; panic responses were certainly not necessary or advisable. Nevertheless there was a shared view that plagiarism was on the increase and this was attributed to two causes. Firstly, there was increased temptation for students to take short cuts to success when more and more were forced into employment to avoid financial problems; secondly, students were inevitably affected by the decline in standards of honesty and probity in wider contemporary society. Computer technology, word-processing and the Internet had resulted in new forms of academic dishonesty which required awareness and vigilance from staff to detect and to combat them.

Many of us see this as more and more of a problem. The University provides a meal-ticket, a hurdle and a rite de passage. Society requires them to go through the process successfully, so the most economical way of getting there is seen as the best. They don't regard this as cheating; we see it as cheating to avoid the effort of scholarship. None of us knows how frequent cheating is; we know less and less about student behaviour. As changes we become increasingly distanced from it. We see them every day and yet about their fundamental attitudes to assessment we haven't a clue. Students are aliens.... [1]

In this context the integrity of examinations should be seen as a powerful advantage. When students were assessed under examination conditions, they could not benefit from collusion and dependence on unacknowledged sources; there was no (or very little) chance of personation; the intellectual ownership of what was produced was not in doubt. Examinations possessed an integrity, shared perhaps by orals, but not by essays, projects, dissertations and portfolios of work; increased diversity of methods in assessment had threatened the integrity

of the assessment process and there might soon be a good argument to retreat from diversity to a renewed reliance on examinations – however unfortunate this might be for those who were unable to produce their best work under examination conditions.

The examination system, especially with anonymous marking of the products, has an integrity that other modes of assessment just cannot match. This may be an old-fashioned view – but its time is coming again. [2]

This quotation points to the second usage of integrity; it recognises that it is not only student dishonesty that can pose a threat to the validity of assessment, so too can assessor bias. (Were there no chance of assessor bias there would be no need for anonymous double marking.) One member of staff emphasised that admitting to the possibility of assessor bias must not be construed as implying malevolence on the part of assessors; the appropriate charge was lack of awareness. Because assessment was always ‘a very inexact science’, then some measure of inter-assessor inconsistency was unavoidable; this, however, did not relieve people of the duty to explore the causes of inconsistency of judgement and to minimise such inconsistency.

We do recognise that the marks a student gets can depend on who does the assessing. I did an exercise once of photocopying three exam scripts and getting everyone to mark them – the range of marks given to each script was worryingly large. I am not sure that we have addressed the question ‘Is your understanding of that descriptor the same as mine?’ with enough seriousness. Are we sure we are operating the same standards here? Are we remembering the criteria and the descriptors as we mark? [5]

Because the fallibility of individual judgements was recognised, procedures were in place to avoid any serious injustice to students. Student products were double, and sometimes triple marked. The class convener sent out profiles of the marks awarded by tutors so that hard, lenient and inconsistent markers could identify themselves. Graduate teaching assistants were carefully prepared before they started marking and their work was monitored.

Considerable effort was expended to ensure that procedures were meticulously followed in the interest of fairness to students and to ensure the ‘integrity’ of the process. One member did, however, feel that in the absence of a shared, full understanding of the meanings to be attached to assessment criteria, as indicated above, all these efforts were directed to convergence rather than accuracy; differing assessments were brought into line, but whether they thus came nearer to what was claimed for them remained uncertain. In the conventional language of assessment, there was more concern for the reliability of assessment than its validity.

(c) Standards drift

There were not many references to standards when interviewees were talking about Level 1 assessment. However, two people did feel that there had been pressure from the faculty ‘to get students through’; there had been comment that the department were not ‘passing’ enough students (i.e. giving grades of D and above) and that philosophy modules were more difficult than those in other subjects. One member felt strongly that not only had standards been pushed down by external forces, but that this unfortunate pressure was likely to get greater. Another had a rather different perspective. There were a lot of students who would fail the Level 1 course if success depended only on the examination; they were ‘pulled through’ by the essay. Before the essay had been given summative weighting, there had been what s/he saw as an ‘unacceptably high failure rate’; in other words a decline in standards was not necessarily a bad thing, if they had been unreasonably high beforehand. There is here a common problem in debate about standards; when a change in an assessment programme results in some diminution in average performance but this takes place across a wider range of

learning outcomes is this better construed as a decrease in standards or merely as a change in standards?

Reference to a downward drift in standards at honours level were made by five of the staff interviewed. The root cause of this was seen as lying in an altered perception of the status of the lower second honours degree. Whereas in the past gaining any 2nd class honours was seen by students as a significant success, it had gradually come to be that a 2.2 was perceived as little better than a fail; only a 1st and a 2.1 were now perceived by students (and those awarding money for postgraduate studies) as 'good' degrees. This has impacted on the marking of student essays and exam answers.

Because one knows that 60 is regarded as a pass-fail borderline, when one reads an essay one tends to think, well this has to get 60 unless I see very solid reasons for marking it down. You tend to work on the presumption that it will get at least 60 unless..... Students are desperate to avoid a 2.2 and you know that they may appeal against being awarded a 2.2. Because one knows that assessment decisions can never be wholly precise, one tends to give students the benefit of the doubt. There is now a cluster of people who end up with an average score of, say, 59.8 to 60.1 and they get a 2.1 when they should really have got a 2.2. [2]

For some people, diversification in assessment methods at Honours had reinforced the downward drift in standards. There was said to be a tendency to inflate the marks given, for example, to project presentations; these took place in an end-of-term atmosphere near the conclusion of the Junior Honours year and some markers were not confident enough about the application of the assessment criteria to mark students down.

I think some markers don't regard the project as very worthwhile – so it seems hardly worth going through the agonising necessary to give something a low mark. An enjoyable theatrical performance with one or two seriously good jokes then scores more highly than it deserves. People can get 70 for the project when they really should get something like 55. [5]

On the other hand, one person was quite sure that the giving of summative weighting to the essay had not contributed to any lowering of standards. An investigation had been carried out which had shown that the inclusion of essay marks had no significant effect on subsequent degree classification; in that particular year there was only one student who would have got a lower class of degree if essay marks had been excluded. There was a tempting argument that the excessive collusion that was possible in the production of essays had resulted in people getting better degrees than they should; there was no sound evidence for this.

(e) Workload

It is a little doubtful whether or not this should be dealt with as a concern at all. Although all interviewees had things to say about the workload generated by assessment, it struck this interviewer forcibly that no interviewee had any serious complaint. The topic is, however, labelled as a concern because this was the one area in which interviewees kept insisting that their views were not typical of philosophy staff as a whole.

I don't think assessment takes up too much of my time; it's not out of hand at all. The people you are talking with are people who are interested in teaching and assessment, not those who see it as a chore which takes them away from their research. [8]

Assessment is not a burden. I do it thoroughly; I would be ashamed of myself if I didn't. My primary enthusiasm has always been teaching, so I never saw assessment - which must be closely related to teaching - as a distraction or as an unreasonable imposition. For some, in the last fifteen years, anything which takes away from one's research is a chore to be performed as quickly as possible. [2]

As a new mode of assessment had been introduced, something else had usually been cut out or reduced. This is somewhat unusual; elsewhere diversification is regularly cited as the most common cause of over-assessment of students and of an unreasonable workload for staff. It seems that it is relatively easy to introduce additional assessments into a course and much more difficult for any existing assessment to be excised. People become attached to existing practice even when aware of its shortcomings. In philosophy there had been a clear saving of time at Level 1; final degree examinations were not formative and there was thus no need to provide feedback for students.

I spend less time on assessment than I once did. We do less marking at Level 1 and spend less time providing comments and feedback. Although of course it is the formative aspect of assessment that is most interesting. [4]

Staff did not feel that assessment took up too much of their time. This does not mean that they devoted little of their time to it; assessment took up far more of their time, they said, than they were ever given credit for in workload calculations. For everyone there were distinct times of the session when the assessment load had serious peaks and for those who had formal duties in relation to the administration of assessment the time spent on this was huge. The production of examination papers, the recording and transmission of marks, the preparation for and follow-up from examiners' meetings and relations with external examiners all took a great deal of time.

Thus assessment took a lot of time, but staff did not think that overall the time was excessive; it probably did not take more time than it had in the past; there was a wish to do more formative assessment, but a general feeling that time was not available for it; staff would probably object to having to spend more time on assessment in the future. Whilst one person said s/he would welcome with open arms any innovation in assessment which saved time, s/he also said that no-one had time available to seek out or to devise such innovations.

G. GENERIC ISSUES

In this penultimate section, five generic issues are addressed. They differ from the concerns identified in the previous section in that the researcher had them in mind before any interviews were carried out. They are issues which it is predicted will be of relevance in all the four subject areas to be studied. There were again no direct questions on these matters in the original schedules. When, during interview people raised these matters, questions of clarification and amplification were asked. In what follows the issue is first defined and then a brief summary of the interview data provided. An apology is offered for the attack of compulsive alliteratitus ('The CAs') in issues two to five.

(a) Assessment: The Big Purposes

In the literature, assessment is sometimes described as having three broad purposes: certification, accountability and learning enhancement. At other times, the first two are conflated and two broad purposes are given: certification/accountability (the control function) and learning enhancement (the growth function).

There was little or no explicit reference to the accountability purpose i.e. that assessment provided evidence that courses were doing what was claimed for them or that staff were doing their job as teachers effectively. Everyone clearly acknowledged the certification and learning enhancement purposes; both were regarded as inescapable.

Assessment is essentially about progress in learning; but it also has to be about giving them a piece of paper which actually means something worthwhile. [4]

Whenever people indicated they would like to see more assessment, or be able to spend more time on assessment, it was always in relation to the learning/growth function. Although there was some concern about a possible downwards drift in standards, people were sure that the certification function was adequately fulfilled; perhaps the leaning function was not.

Assessment is not interesting when it's about certification. I have no interest in putting people in pigeon holes to satisfy the requirements of the outside world, except to do it in a fair way that does justice to students. [1]

Some of the requirements of assessment come from people who have experience only of manufacturing jam and these people don't understand academic life and its purposes and standards. We can never divorce assessment from student learning – that would be disastrous. [7]

(b) Constructive alignment

Constructive alignment is a concept which has appeared only recently in texts on assessment. It refers to that situation in which the aims and objectives of a course (perhaps described as statements of learning outcomes), the methods adopted for teaching the course and the ways in which learning is assessed are all planned together and related to each other so that each guides and reinforces the others. Outcomes, methods and assessment then co-exist in a positive, productive relationship i.e. they are constructively aligned.

This type of discourse was only employed by two of those interviewed. It is probably significant that these two appeared to be those with the greatest volume of experience of quality assurance procedures and in assessment deliberations at faculty or institutional level.

I think perhaps that rather few of us are aware of the importance of aligning assessment with the specification of aims and objectives; you don't have that awareness if you are only involved in the department or the philosophy community. I see this as important because of the roles I've had outwith departmental teaching.[6]

In what was said by others there was detected a rather different form of alignment, what we might call epistemological alignment. The justification for assessing students by particular methods, or for changing methods, lay within the perceived nature of philosophy as a discipline or form of knowledge. This was particularly obvious in relation to two assessment methods, essay writing and project presentations. The importance attached to essay writing was appropriate because it was an accepted and necessary part of being a philosopher that one developed coherent arguments in extended written form that remained available for thoughtful scrutiny by other members of the academic community. However, it was also intrinsic to the philosophical pursuit that one was able 'to talk philosophy'; generating philosophical talk and then pursuing an argument in response to the immediate comments made in discussion provided a subject-based defence for the introduction of the project-presentations. If philosophy is construed as 'a series of edifying conversations' then students

must converse in situations where their edifying quality is a matter of interest. This argument also underpinned the wish to see more of students in individual and small-group tutorials.

This type of alignment-with-subject argument is plausible in relation to the two assessment methods described. It is somewhat dubious as a defence of the continuing importance attached to degree examinations; it does tend to be used in almost all subjects irrespective of their nature. Such examinations are now commonly criticised for their lack of constructive alignment with intended learning outcomes; the integrity of such examinations seems a better argument for their retention than does epistemological alignment.

(c) Communicating about Assessment

There is a strong case made in some assessment literature that assessment only reaches high levels of reliability when the assessors involved are members of a regularly inter-communicating group. There is a naïve belief that reliability in assessment can be achieved through the clear explication of assessment criteria. People are then disappointed when a conscientious exercise in clarifying and agreeing criteria does not remove inter-assessor inconsistencies of judgement. Clear assessment criteria are a necessary but not sufficient condition for reliability; what is needed in addition is that assessors have a shared understanding of what these criteria actually mean. Such shared understandings commonly come about by assessors discussing and arguing amongst themselves about how the criteria are to be applied. (The creation of mark schemes and model answers also helps to attach meaning to criteria.)

There was material in all the interviews that related to the theme of communication about assessment. Two matters were very clear. Firstly, there was an impressive amount of collaboration over the construction of assessment instruments, particularly exam questions. Course teams did communicate with each other round a table, but perhaps more usual was the circulation of drafts for comment, increasingly done by e-mail. Secondly, there was quite regular discussion about individual student scripts and dissertations: 'there have been serious ding-dongs about the worth of a specific script' and 'there are quite frequently frightful arguments about controversial dissertations'.

There was, however, a majority view that, apart from the two instances above, there was little departmental discussion about assessment issues. This was not surprising, and certainly not reprehensible, given that staff had known each other for a long time ('We've talked together a great deal over the years') and were familiar with each other's thinking. There had been plenty of time and opportunity to develop shared understandings about assessment criteria and how individual members of staff operated them. The amount of within-group understanding was seen as being particularly obvious (a) when a new external examiner was appointed ('it can take time to socialise them into our ways of thinking') and (b) when new graduate teaching assistants were being inducted into marking procedures ('we talk to them a lot and watch them carefully until they become familiar with our ways').

Although everyone said something related to communication, two people spoke at some length on issues of particular interest to them. The first stressed that the amount of discussion at examiners' meetings had decreased. This was because summary assessment decisions about final grades or degree classification were now made by numerical aggregation procedures rather than by individual tutors contributing what they knew of students and their work and then reaching a consensus of professional judgement through discussion. This changed process led to decisions which might be easier to defend but were not necessarily wiser. An inevitable by-product of the change was less explication of, and discussion about, the meaning of the assessment criteria which were being operated.

In these meetings you now get translation of letters into numbers, averaging, re-averaging and re-translation back to a letter or a degree class. It is essentially a formulaic process and it's difficult to argue about the verdicts reached. In the past you could focus discussion on key cases, the best and the border-liners and the strange. That's impossible now; candidates are not recognised as people. [3]

The second person had wide-ranging concerns relating to communication about assessment within the department. What s/he said on the topic is now given verbatim, without comment.

There is really remarkably little discussion of assessment during the year. 'When you say "poor" on this piece of student work, why are you saying it? In what ways is it poor? What are your expectations of "poor" and "not poor"? We don't do this as any kind of routine – but only in response to a particular student decision or an appeal. There are procedures for settling particular issues, but these don't raise the wider questions.

Discussion doesn't come out of double marking as you might think. In nearly every case I think the second marker just acquiesces, defers to the first judgement. People tend to say 'You know more about this topic than I do' and this is just too easy. They ought to say 'Perhaps you know more about the content – but are you so familiar with it that you don't keep the marking criteria in mind while you're marking it?'

Of course pressure of time is a factor. In the past fifteen years or so there has been a decrease in emphasis on teaching and assessment; but it is not just this. I don't think there is any great enthusiasm to look at this kind of thing. If a staff meeting was suggested to talk about assessment, there would be groans and everyone would say that there is more important business to worry about. People shy away from it, perhaps thinking there is nothing much that anyone can do about it and that it doesn't really matter that much anyway.

We do too little talking about assessment and I put it on the back-burner as cheerfully as everyone else. [5]

(d) Criterion-referenced assessment

The distinction between norm-referenced assessment (spreading candidates out to make judgements about their relative standing) and criterion-referenced assessment (judging candidates' absolute standing against predetermined criteria) is a familiar one of long-standing. It is important to ask what these things are. They probably represent polar extreme schools of thinking about assessment or even ideologies to which people claim allegiance in an uncritical way. It is doubtful if they can ever be used accurately to describe any actual assessment scheme or strategy. In practice, pure forms of norm-referenced and criterion-referenced assessment do not exist.

In the philosophy interviews, there were regular references to assessment criteria and to the desirability of making criteria explicit. One person spoke of the value of having criteria-related grade descriptors: 'When numerical grading schemes keep changing, it is very useful to have grade descriptors to keep in mind'. Apart from these points there were no explicit comments about criterion-referenced and norm-referenced models of assessment or about the difficulties of resolving the tension between the two in an actual assessment scheme.

(e) Constrained autonomy

The idea of any single academic having complete freedom to do what s/he thought best in assessing students is, and always has been, absurd; individual autonomy is always constrained. (In philosophy it was noted that it had become more constrained as individual members worked more in course teams.) The idea of a group of academic staff in one subject area having freedom with respect to how 'their' students are assessed is not self-evidently ridiculous. However, even here clear constraints are recognised. Two examples make this obvious. Firstly, if students may pursue a joint honours degree in, say, philosophy and history, then the arts faculty must have some policy which ensures that two assessment systems are comparable. Secondly, students will be awarded a degree not by the philosophy department, but by the university; thus the university clearly has to exercise its legitimate authority over what is done in its name. This manifestly includes how results are recorded, aggregated and conveyed to a central registry, but it also includes aspects of quality control.

The existence of such constraints on individuals and the department were seen as entirely proper and reasonable, even when details of how they emerged and were exercised were sometimes criticised. However, people did have worries about how their autonomy might be increasingly constrained in the future.

(i) Constraints might extend into new areas of activity.

There was an acceptance, at times somewhat qualified, that University policy legitimately directed activities in administrative, bureaucratic and some certification/accountability aspects of assessment. Those interviewed would react negatively were The University to attempt to prescribe what within philosophy they should assess and how they should assess it. The University was perceived by a few as getting dangerously near such prescription.

(ii) Constraints might increasingly come from supra-institutional sources.

There was a decidedly adverse reaction against the influence of national policy-generating bodies, vaguely 'the government' and specifically 'the QAA'. They were perceived as increasingly failing to appreciate the values of academic life and the nature of subjects like philosophy.

External influence was not seen as malign when it came from people with the appropriate subject-based expertise and credentials. For example, the external examining system was seen as a constraint, but an acceptable one because the examiners were themselves philosophers. This is somewhat similar to the view of one person who would react against input from the Institute of Learning and Teaching on how to assess students, but would welcome guidance from the appropriate subject centre of the Learning and Teaching Support Network.

(iii) Constraints might be imposed in a more directive way.

Policy does not have the status of law and it does not unambiguously prescribe what has to be done. Policy can often be re-constructed rather than merely implemented. External constraints in the past have often allowed some room for manoeuvre, some scope for interpretation (although this had not applied in the case of modularisation). There was an apprehension amongst most of those interviewed that 'management' and 'policy' were likely to become more intrusive and more directive in relation to assessment.

I do think a lot of assessment requirements are just imposed on us, because of the structure and character of the institution as a whole and the way it is increasingly being run. You just have to do your best within the framework that others provide.

Some impositions are of course better than others and sometimes it is not hard to make them fit with subject requirements. Often we have wasted a lot of time trying to tailor what we do, to fit in with what seems to be externally-imposed policy. Can we continue to do this, I wonder? [7]

H. CONCLUDING REMARKS

In the opening paragraph, three research questions were identified. Why do you assess students in the ways you do? When assessment practice changes, why does it change? What assessment issues emerge as important? Answers to these questions summarising the views of the interviewees will now be attempted; brevity produces some distortion.

- Philosophy students are assessed in the ways described in Section B and these ways were not seen as very different from those used over a long period. There had been no compelling reasons for radical change; the assessment systems worked well.
- Level 1 courses had been re-structured; changes in assessment were introduced as part of the general process of course planning.
- There had been changes in Honours Level assessment; these had given more emphasis to continuous summative assessment, to diversity in assessment methods and to the assessment of presentation skills. Influential factors were institution-wide expectations and QAA guidelines on assessment.
- Philosophy was seen as a distinctive subject; assessment practice and assessment changes, at Honours Level particularly, were seen as being aligned with the nature of the subject.
- So far, external policy requirements had not dictated major changes in assessment practice; such requirements were expected to become more influential in future.
- Those philosophy staff interviewed were highly conscientious about assessment and did not complain about the considerable workload involved.
- There was a general view that more time should be spent on formative assessment and on that informal assessment which cannot be separated from effective teaching.
- There was a minority view (a) that the academic philosophy community as a whole and the department should be more innovative in its thinking about assessment and (b) that philosophy staff in the University should spend more time communicating with each other about the reliability of their assessments and the contribution assessment can make to enhanced student learning.

ASSESSMENT IN HIGHER EDUCATION

Subject Area III: Medicine

Descriptive Report: Version 2

A. SETTING THE SCENE**(a) The Research Purposes**

The research aims to illuminate the underlying rationale for current practice in assessing student learning; it aims for description and explanation only. The researcher has no role as an evaluator, although the evaluations of others are described. This phase of the research is guided by three general research questions: Why do we assess students in the ways we do? When assessment practice changes, why does it change? What issues in assessment emerge as important?

(b) The Course

The focus of the research was assessment of student learning within the five-year MB.ChB. Degree programme. A new curriculum came into operation in 1996; thus when the interviews to be described were carried out the first students following the new curriculum were about two-thirds of the way through their final year. The final assessments had been planned, but no students had experienced them.

It is assumed that all readers of this report are familiar with the aims of the new curriculum, the educational principles underpinning it and the variety of ways in which student learning experiences are structured. What follows is a brief reminder only of the context in which assessment occurs.

- (a) The educational strategy adopted for core learning throughout the programme is Problem Based Learning (PBL). In years 1 to 3, students work in small groups (normally of eight people) each with a facilitator and each following a general procedure (The Glasgow Steps) designed to encourage independent learning. Each five-week PBL block has a general theme and is organised around five reality-based scenarios (two PBL sessions per week). PBL sessions are supplemented by Fixed Resource Sessions involving laboratory work, skills training sessions, visits and so forth. There are six PBL blocks in year 1, five in year 2 and four in year 3. This student-centred learning approach is continued throughout years 4 and 5 but with different emphases and methods of operation to reflect the students' growing maturity and the increasingly clinical orientation.
- (b) Vocational studies/clinical practice are pursued in years 1 to 3 in parallel with PBL sessions.
- (c) There is no rigid distinction between pre-clinical and clinical phases. The majority of the first two years is university-based; the majority of years 3, 4 and 5 is hospital/community-based. In years 4 and 5 there is continuous rotation through seven clinical attachments each lasting either five or eight weeks.
- (d) Also within the programme are seven five-week blocks in each of which students take a Special Study Module (SSM). Students choose from an extensive menu of options (or may propose and design a module for themselves); these may deepen or extend

their medical understanding or allow study of some subject conventionally outwith medicine, but of educational relevance.

- (e) If students wish, and are then selected, they may pursue an intercalated degree after year 3. They also add a short period of elective study between years 3 and 4 (the Junior Elective) and between years 4 and 5 (the Senior Elective).
- (f) The degree programme has been planned to satisfy the recommendations contained in the report from the General Medical Council (1993): *Tomorrow's Doctors*.

There are about 240 students accepted to study medicine each year, although on occasion the number has been rather higher than this. In other words when the new curriculum is fully operational the new assessment arrangements will have to cope with a total of about 1,200 students. (The final assessments this year cover a smaller number of 'new curriculum' students because some are pursuing an intercalated degree; in addition there are, of course, a roughly equivalent number of 'old' curriculum students returning to the clinical years after pursuing an intercalated degree.)

(c) The People Interviewed

General approval for the research was obtained from the Associate Dean for Education. He was then asked to suggest the names of members of staff whom it was recommended should be approached for interview. After consultation with the Curriculum Development Officer, twelve names were provided.

The twelve people were approached by letter at the beginning of January 2001; the letter covered the purposes of the research, the main topic areas for interview and an indication of the time that might be required. Enclosed was a protocol making explicit the 'rules of engagement' and covering aspects of approval and anonymity. The letter was followed up by an e-mail a week later. The researcher was apprehensive at this time; he had been well warned that all of the people approached were exceptionally busy and might well not be willing to find the time to co-operate in the study. Within forty-eight hours of the e-mail, eleven positive responses had been received and dates and times arranged. One person declined to take part; the reason given was exceptionally interesting and is considered later under emerging issues (Section Gc).

The eleven people interviewed were as follows: the Associate Dean for Education, the Curriculum Development Officer, the Clerk to the Faculty of Medicine, the Year 1 Co-ordinator, the Year 2 Co-ordinator, the Year 3 Co-ordinator, the Co-ordinator for Years 4 and 5, two Facilitators (with special duties relating to assessment), the Special Study Modules Co-ordinator and a Vocational Studies Senior Lecturer with responsibility for assessment in that area.

It is rather difficult to say how many people form the population of 'those who "teach" within the medical curriculum'. There are ten PBL facilitators; they all have temporary non-clinical academic-related contracts, either full or part-time. Also located within the Medical Education Unit is the Curriculum Development Officer. There are about 300 lecturers, senior lecturers, readers and professors from the more than thirty departments within clinical medicine alone. The faculty of medicine also buys-in the services of staff from the Institute of Biomedical and Life Sciences. In addition it has been calculated that there are over 1,400 National Health Service staff who are involved in one way or another with the education of undergraduate medical students. It is **not** claimed that the eleven people interviewed in this research form any kind of representative sample of this huge population of 'those who teach within the medical curriculum'. Rather they form a purposive sample, chosen because they

satisfy the criterion of being well-qualified, through a variety of types of experience and involvement, to speak on the topics covered in the interview. There is no intention to make any empirical generalisations beyond this group.

With each person's permission, all interviews were recorded on audio-tape. Interviews lasted on average for 56 minutes (excluding points in preamble and farewells) and ranged from 45 to 70 minutes. It is remarkable that people facing heavy and conflicting demands were willing to give so very generously of their time.

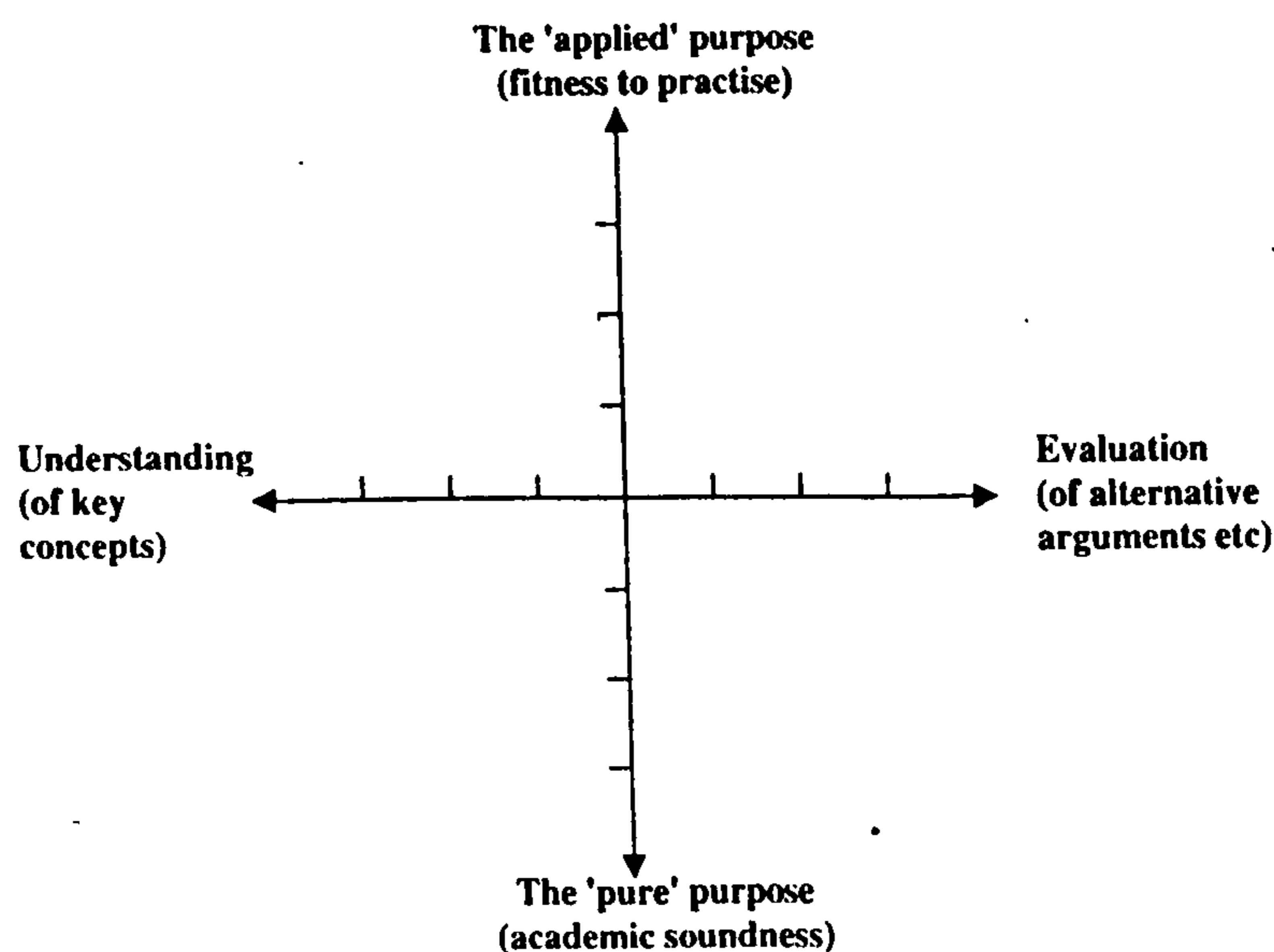
(d) Medicine as a Subject Area

In choosing medicine as one of the subject areas to be explored, there was an assumption that it made sense to treat medicine as one subject area. This assumption was confirmed in interviews as being justifiable for the new curriculum.

It's certainly not silly to see it as one subject area, although it might have been in the past. It's a large collection of specialisms united by a common professional concern. What did someone call that – a field of knowledge?

[1]

(e) 'The Four Quadrants' Where is medicine?



Interviewees were given the diagram above; it was introduced like this. 'It is said that subject areas are differently located on this diagram. The axes represent what is emphasised in assessment near the end of the course. Please do two things, (a) comment on the framework itself and (b) whatever your reservations, place medicine where you think most appropriate.'

The first point to be made is that, whatever reservations people had about this simple diagram, almost everyone had a lot to say about it. When invited to assess the strengths and weaknesses of this theoretical framework, these medical teachers responded with relish and apparent enjoyment. In general the framework was viewed favourably: 'this diagram makes sense' and

'this is a good grid – it makes me think'. There was a tendency for people to talk about what the medical curriculum **should** be like, rather than what was actually emphasised in final year assessments; 'I'm talking here about my hopes rather than what is'. It does seem likely that in this new curriculum there is a genuinely close correspondence between what the curriculum is claimed to be like and what actually gets assessed i.e. between the rhetoric of course purposes and the reality of assessment practice. Confirmation of this would require not only analysis of what is sought through assessment instruments and procedures, but also scrutiny of what gets rewarded in student responses. This goes well beyond the present study.

The nature of the vertical dimension was seen as much less problematic than the horizontal dimension. 'Fitness to practise' must be based upon 'academic soundness'; however, the inescapable central concern at the end of the five year programme must be whether or not students could be judged as fit and safe to practise (albeit under supervision); this provided the dominant emphasis in the end of programme assessment. When considering the horizontal axis, staff were hesitant about accepting the poles as genuine opposites. For example, evaluating alternative arguments/procedures was central to the making of differential diagnoses and judging the merits of possible treatment plans; however such evaluations had to be based on good understanding of certain core concepts; final year students must be able to demonstrate and explicate their understanding i.e. it must be overtly assessed. Several of those interviewed identified the two ends of the horizontal axis as referring to physical and social sciences i.e. they saw this as being a 'hard-to-soft' dimension.

Ten of the eleven people located medicine on the diagram. Two people said that assessment was necessarily about everything on the framework and thus the only sensible place to locate medicine was in the centre. The remaining eight were reluctant to place medicine at **one** location on the diagram; they insisted on two points, one for the previous curriculum and one for the new curriculum. The average position given for medicine in the old curriculum was in the upper left quadrant at (-2.0, +1.5); for the new curriculum, medicine was located just into the upper right quadrant at (+0.5, +2.0). Medicine was chosen as a subject area for the research in the expectation that medical teachers would locate it in the upper-left quadrant; this was indeed the result for the old curriculum. The shift perceived for assessment in the new curriculum up the vertical axis is probably not sufficiently marked to be significant; the perceived shift from left to right is exceptionally interesting. There seemed to be three explanations offered: (a) there is less emphasis in the new curriculum on the physical science knowledge bases of medicine and more on social and ethical aspects; (b) there is now more emphasis on making use of knowledge than on possessing it and (c) there is more emphasis on informed critique of alternative theories, diagnoses and treatment strategies.

B. STUDENT ASSESSMENT: THE PRESENT

(a) Clarification of terms and acronyms

Before describing how students are assessed in the new undergraduate medical curriculum, it may be useful to clarify some of the terms and acronyms used. Some of these may be unfamiliar to non-medical readers; some are applied to a local modification of a generally familiar practice. That this preliminary clarification seems necessary is interesting in itself; there was no need for such an exercise in reporting on assessment in the first two subject areas. It seemed a safe assumption there that everyone would be familiar with examinations, essays and objective testing. There has been a lot of innovation in medical assessment and within this curriculum there is considerable diversity within assessment practice.

MILE **Medical Independent Learning Examination**

This examination mimics much of the normal process of problem-based learning. As the name clearly suggests it aims to assess each student's ability to learn on his or her own; it is an assessment of the achievement of outcomes concerned with learning process skills and although obviously not content-free, it aims to be content independent. The original conception was Australian, but it was adapted for the Glasgow situation and the version used is probably unique to Glasgow.

Each student is given a clinically-based scenario; for any one year group there are three such scenarios generated and each is presented to about eighty students. Students have one day to define problems, to formulate the questions which they wish to answer, to research answers to these questions and to write up the process and their answers to questions identified. Students can work in groups or on their own as they choose. A group of six staff tackle the scenario themselves, then mark a few student reports and develop a mark scheme. Each scenario is then marked according to the same five assessment criteria and with the help of the scenario-specific mark scheme. The group of six assessors spends a whole day together doing the marking, discussing the meaning of criteria and double-marking low and high-scoring scripts and problematic ones. This process is recognised as time-consuming, but the social/communicative aspects of the assessment procedure and its time-limited nature are welcomed.

OSCE **Objective Structured Clinical Examination**

This assessment process requires students to circulate round a number of stations (typically about fifteen) organised in a circuit. At each station they are required to demonstrate some aspect of clinical performance. This could be the carrying out of a specific clinical procedure (e.g. measuring blood pressure) or exercising some communication skill (e.g. explaining a diagnosis to a patient). In these examples student performance would be assessed by a trained observer against clearly explicated criteria. At some other stations, a written product from the student might be assessed.

OSCEs in one form or another are now widely used in medical schools throughout the world; it is interesting, however, that early development work in the technique was carried out in Glasgow and then Dundee. They are recognised as being challenging to design and time-consuming to set-up and administer; however, relatively little time is required of assessors after the event. The adjective 'objective' should probably be interpreted as 'relatively non-subjective'. Because of the careful analysis of clinically-related tasks, the availability of clear assessment criteria and the standardisation of patient behaviour that is possible, such assessment claims greater reliability than other forms of clinical assessment without over-much loss of authenticity and validity.

MEQs **Modified Essay Questions**

Students are given an account of some clinical condition or situation and are required to write a set of relatively short, focused and highly structured answers dealing with different aspects of the topic. The clinical scenario integrates a number of disciplinary perspectives. The main purpose of the assessment is to get at student understanding of the component knowledge bases.

The MEQ is presented to students as one booklet, with each page containing carefully structured questions on a different aspect of the topic; students provide their answers within the booklet. This means that after the examination, each page can be separated and a collection of all the student responses on one aspect sent to a marker with specialised knowledge of that aspect.

This should be seen as constructed-response assessment, but the degree of structure involved should mean more reliability in marking than is normal for traditional essay type questions.

EMIs (or EMSQs) Extended-Matching Items (or Extended Matching Set Questions)

EMIs are a form of objective or selected-response testing. Although standard Multiple Choice Questions (MCQs) are still common in medical assessment (particularly in postgraduate selection for Royal Colleges although there is movement away from MCQs here also), it has been decided for this undergraduate medical curriculum that EMIs avoid some of the limitations of standard MCQs in that they make it somewhat easier to assess higher cognitive abilities and relationships between otherwise separated bits of knowledge.

Extended Matching Items can be defined as multiple choice items organised into sets that use a list of options for responding to all items in the set. A typical item will have a unifying theme, a list of up to say a dozen options, a lead-in (for example, 'Select the most likely diagnosis') and a number of stems. Each stem will require the student to select one or more options from the list. Responses to these questions can be marked by computer.

MLC Modified Long Case

This form of clinical assessment has been developed from the familiar Long Clinical Case assessment. As practised in this curriculum it has four phases: (a) the student meets a patient, takes a short medical history and carries out a physical examination; (b) the student has twenty minutes on their own to think and to consult standard texts; (c) the student has a fifteen minute Viva on the case and (d) the student returns to the patient to explain future case management. Two examiners (one from the main specialism involved in the case and one from a quite different specialism) observe the student in phases (a) and (d) and examine him or her orally in phase (c).

LCP Longitudinal Care Project

Strictly this is not a method of assessment, but a student experience which is assessed through a written assignment or project report. The student is required to describe, and to reflect critically upon, clinical practice within the community gained through attachment to General Practice.

The Assessment Grid

It is intended that much of the assessment throughout the five years of the course be related to one single underpinning pro-forma. This assessment grid has four main domains: The Doctor as Professional, Approach to Professional Work, Clinical Competence and Integration. The first three domains are subdivided into major areas; for example, clinical competence involves history-taking, communication skills, clinical examination skills, clinical judgement and practical procedures. For each area in each domain there are verbal descriptors of three levels of performance – highly satisfactory, satisfactory and unsatisfactory. For example, to be highly satisfactory in Area D 'Knowledge', the student must be judged 'to have an extensive and appropriate knowledge base and be able to apply it to clinical problems and to be eager to extend their knowledge'. In the first three years of the course the first two domains have priority; by year 5 all four domains are necessarily involved.

The above do not cover all the forms of assessment in the course. There are other techniques which probably do not need clarification; these include Writing Short Notes, Data Interpretation Exercises and Portfolios.

(b) How are students of medicine currently assessed?

A relatively brief description of assessment in the new curriculum should now be possible. This has been constructed from the interviewees' accounts of how students are assessed and not from course documentation. The writer has been unable to find a summary account of the assessment strategy for the new curriculum and would recommend that such an account be written and made readily available, primarily for students but also (a) for those members of staff who are very familiar with assessment in one part of the course but a little uncertain as to what happens elsewhere and (b) for those outwith Glasgow who find current developments in medical education in Glasgow a source of great interest.

Year One

(a) Coursework Assessment

Students are required to submit four end-of-block assignments; each is marked on a 20-point scale. Coursework as a whole carries a 20% weighting in the final grading for the year.

(b) Degree examinations

There are two written papers, each of two hours duration. The first paper contains one substantial MEQ and a number of short notes; students are advised to divide their time equally between the two components. The second paper contains another MEQ and then an interpretation exercise. (The latter might be based on a 'populist' medical article or a journal editorial.) The degree examinations count for 80% in determining the final grade for the year.

(c) MILE

Students sit the Medical Independent Learning Examination as described above. This examination must be passed, but the result of it does not affect a student's end-of-year grading.

(d) Other assessments

At the end of each scenario, of which there are five in each five-week block, students are asked to consider how well they have functioned as a group. Did they set themselves appropriate questions? Did they go about answering them in effective ways?

At the end of each block, of which there are six in the year, the facilitator leads a discussion which aims to be a critical evaluation of the group's progress and then provides an opportunity for each student to have a short private discussion about their individual progress. It was stated more than once that PBL facilitators had no role as assessors: 'it is their job to assist not to assess'. It would be accurate to say that facilitators made no formal input to the summative assessment of students; it is clear, however, that they have a highly significant role within that informal assessment which cannot be separated from learning.

Year Two

(a) Coursework assessment

As for Year One, but five end-of-block assignments required. 20% weighting.

(b) Degree examinations

Similar to Year One. Original provision was for two three-hour papers; this has been reduced to two two-and-a-half hour papers. 80% weighting.

(c) OSCE

There is an Objective Structured Clinical Examination in Year Two. This must be passed, but it does not affect the end-of-year grading.

(d) Other assessments

These are as described for Year One.

There is one SSM in Year Two – see the separate section.

Year Three

(a) Coursework assessment

In this year, the coursework assessment is based on the LCP (see above) 20% weighting.

(b) Degree examinations

These are similar to those in Years One and Two; again they contain MEQs, Short Notes and an Interpretation question.

(c) OSCE

This is similar to the Year Two OSCE, but somewhat longer and more demanding. There are twenty-five stations; all students sit fifteen of these and can gain exemption from the other ten if their performance is adequate. Satisfactory performance is again essential for progress, and also has some influence on future decisions about Honours/Commendation.

(d) Other assessments

In theory these were the same as in the first two years; there was some doubt expressed as to whether facilitators had the time to carry them out, or indeed whether such assessments were still as necessary as in the earlier years.

There are two SSMs in Year three – see the separate section.

Years Four and Five

In the previous curriculum there was a clear separation at the end of Year Two between the pre-clinical and clinical phases of the course. (One interviewee referred to pre-cynical and cynical phases.) This meant that assessment in the first two years was predominantly of knowledge and understanding and in the last three of clinical capability. In the new curriculum there is a clinical/practical emphasis right from the beginning and an insistence on the underpinning knowledge bases right to the end; this is reflected in the assessment scheme. However, the nature of the two year rotation through clinical blocks has tended to create a separation at the end of Year Three which was not the original intention of the course designers.

(a) Assessment during clinical attachments

Over the two years, students rotate around seven clinical attachments (two medical, two surgical, obstetrics and gynaecology, paediatrics and psychiatry) and one general practice placement. Each of these lasts either for five or eight weeks. Each student is assessed in each attachment by an individual with whom they have had a fair amount of contact (the educational supervisor, usually someone with the status of a consultant or principal in general practice). Half-way through the attachment, the student completes an assessment grid (i.e. self-assessment) and discusses it with their educational supervisor. At the end of the period, the student again meets with the supervisor and the latter grades the student on each aspect and then provides one overall summative grading of either Highly Satisfactory, Satisfactory or Unsatisfactory. Supervisors may, if they wish, choose to set an end-of-block examination to inform their final decision; this form of assessment dominated in the past.

Any student judged Unsatisfactory on one attachment in Year Four can repeat that block over the summer vacation; any student Unsatisfactory in two blocks would not be permitted to proceed. Students have to be at least Satisfactory in all attachments to complete the degree programme as a whole (and thus to be allowed to take Finals).

(b) Final written examinations

These are held around Easter of Year Five. There are two written papers assessing the quality of students' knowledge and understanding. These are based mainly on MEQs and on EMIs, but each contain a short note question as well. The latter tends to focus on some ethical or attitudinal issue.

As in Years One to Three the marking of these written papers results in a numerical score. There is thus generated a range of numbers and examiners then use their professional judgement (in consultation with external examiners) to decide where the boundaries between Pass and Fail should be.

Year Five students sit a 'mock' version of the final written papers in November of the previous calendar year; students in Year Four can sit a 'mock' final exam at the same time as the Year Five final and mark themselves on this. (It was noticeable that some interviewees referred to this as formative assessment, but that others used the word in different ways. This point is discussed further later.)

(c) Final Clinical Examinations

These two assessments are held during Term 3 of Year Five. The first is an MLC as described above; the second is an OSCE of twenty stations. Performance of at least a satisfactory standard is required in both.

At the same time students have to present a portfolio of cases from their clinical attachments. They elect to write-up their twenty 'best' cases; examiners select two of these for scrutiny; these are taken from disciplines other than the one focused upon in the MLC.

(d) Other assessments

Students are required to take four SSMs during Years Four and Five. See the next section.

SSMs Special Study Modules

Students take a total of seven SSMs spread over Year Two to Year Five; they are required to gain at least a Satisfactory grading in at least six of these.

The planning, provision, administration and assessment of the Special Study Modules is a huge task. A very extensive menu of options is required to provide opportunities for 240 students to study seven SSMs each. Some of the Year Two modules attract up to 60 people each; some of the Year Five modules run on a one-to-one basis. The SSMs vary widely in nature from 'conventional' taught courses to student-proposed autonomous learning projects and thus the assessment methods must also be varied if they are to be appropriate to the module purposes. On the other hand there has to be some comparability across the very different modules; there has to be some way of reassuring a variety of interested parties (including external examiners) that all SSMs are, at the least, comparable in assessment load and rigour.

Any person designing an SSM has to make the assessment methods and their relative importance entirely explicit. They are encouraged to be creative in their teaching and assessment procedures; however, whatever assessment they propose they must use generic assessment criteria which apply to all modules. In a detailed procedural Guide made available to staff, verbal descriptors of the required standard of performance in relation to each criterion are provided. On each criterion, students are to be graded at one of five different levels (Fail, Unsatisfactory, Satisfactory, Highly Satisfactory and Excellent); there is then a procedure for aggregating these grades to produce a final grade for the SSM (which has to be Merit, Pass or Fail).

A number of other rules for the conduct of SSM assessment are also provided; as just one example, 50% of the assessment within any SSM has to result in a tangible product which is open to inspection by external examiners. Monitoring the appropriateness of assessment methods and the effectiveness with which they are implemented is the province of the SSM Co-ordinator. ('I am the gatekeeper who determines what is and is not acceptable assessment.')

A rapid scrutiny of the menu of SSMs showed evidence of great creativity and innovation in course planning; the assessment strategies did, however, seem to place great reliance on somewhat conventional written reports. This may be inevitable; the SSMs may be thought by some to be more 'risky' than other parts of the curriculum. They then need reassurance, via assessment arrangements that are traditionally perceived as 'rigorous', that there will be no drop in standards.

(c) What is seen as 'good' in the current assessment of students?

This section aims to capture what interviewees said in response to the question: 'What in your view is good (i.e. pleases you) about how students are assessed within the new curriculum?' The intention is to summarise the judgements that people made of assessment procedures; the researcher is not acting as an evaluator, but as a recorder of other people's evaluations. Before embarking on this, three general points seem necessary.

Firstly, all interviewees (with one possible exception) seemed to find this a perfectly reasonable question; they had no difficulty answering it and some answered at length. This was not the case in the first two subject areas studied; in those people appeared to think the question a somewhat unusual one and one they had never asked themselves. Perhaps because

of the newness of the medical curriculum, people had quite recently been taking part in the design of assessments, in discussing the production of new types of assessment instrument and in reflecting upon their assessment experiences. For whatever reason, these interviewees appeared genuinely interested in assessment and were happy to talk about the positive features of the scheme they were busy implementing. (Whether this lively interest in assessment extends beyond the people interviewed has to be a matter of speculation.) However, one response was distinctively different from the others.

What pleases me about our assessment of students? Not a lot, I can tell you. A hell of a lot is non-pleasing.... [4]

The second general point is that most interviewees stressed the newness and developing nature of the assessment scheme. This was especially so for people talking about final assessments in Year Five; assessment had been planned but had not occurred at the time of interview for the first cohort of students on the new curriculum.

Here I am really talking about what we hope will be good; some of it hasn't happened yet, so we don't have evidence of how well it works. [2]

Enthusiasm in talking about positive features of assessment practice was tempered by a proper awareness that no perfect assessment scheme has yet been devised anywhere and that this particular scheme could be developed and improved. The view conveyed was thus one of justifiable pride in what had been achieved, rather than easy complacency about the future.

We are doing well, but not as well as we could do. It's not right, but that would be a silly thing to say about any assessment system. But we have quite definitely made big strides in the right direction. [1]

I think what has been established is the basis for a lot of good practice in assessment. But things are not as good as they should be. Perhaps we have not put in as much effort into assessment as we should have done..... [3]

The third preliminary point is about the analytical character, frankness and open-ness of the interviewees. It would have been understandable if people had produced some fluent in-group rhetoric designed to convince this 'outsider' researcher (and perhaps even themselves) that everything in the medical garden was rosy. Only on two brief occasions within the eleven lengthy interviews did this interviewer feel that he was there to be persuaded rather than enlightened.

Interviewee responses can be placed into six categories. These categories were formed only when at least five people provided broadly similar opinions categorisable under that heading. The six are: the 'workability' of the scheme; student and staff reactions; individual methods; diversity and balance of methods; integrative emphasis and coherence.

(i) The assessment scheme works.

This may seem rather a minimal justification for pleasure, perhaps a necessary but wholly insufficient condition for the effectiveness of any assessment strategy. However, it has to be remembered that these medical staff had been involved in mounting a radically new course, only one component of which was a significantly different assessment scheme. Of course there had been real fears and uncertainties, a genuine anxiety that the whole thing might just not work. The evidence points to the conclusion that it has worked and is working; this is a good justification for relief and satisfaction.

There have been minor crises and administrative glitches, but the planned assessment scheme has been put efficiently into action, most of the intended assessments have taken place, results have been processed and published and decisions on student progress have been taken and implemented. Systems and people have been under strain, but they have not collapsed.

The assessment system was said 'to work' in an even more important sense than the administrative. It appeared to be working in the sense that it had identified as succeeding those students who ought to be so identified, i.e. the assessments seemed to be valid as well as manageable. Although this last point was made by six people, it should be noted that one of them did have a reservation: s/he thought that although assessment did identify the 'best' people, those students for whom medicine was the wrong course choice were not being 'weeded-out' as effectively as in the past. This would seem to be one aspect of the validity and fairness of the assessment system which should be the subject of review and perhaps of research.

(ii) Student and staff reactions have been positive.

Assessment is not commonly welcomed by students as an enjoyable experience and no doubt medical students have at times found assessment in this new medical course both burdensome and stressful. It is thus remarkable that staff identified the pleasure they had experienced from positive feedback from students.

I am pleased that students cope with the assessments so well; they cope with the whole curriculum and assessment much better than anyone dared to hope. When I say 'cope' there is evidence of this in the quality of the products we assess. But it is also that students come up to me after exams and say 'thank you – that was great'. I do find this very surprising. It even happens with students who fail and have to re-sit; they say to me 'Now I realise what I'm supposed to be doing'. Realisation of their strengths and weaknesses does come about through appropriate assessment. [5]

Staff reaction was also said to have been positive.

Staff are pleased to discover that they can generate new types of questions and assessment instruments that do actually work. Most of the sceptics have been converted into co-operation. [1]

It is worth noting that the staff who have had most to do with assessment in the new curriculum are not always those who were most involved in the previous curriculum. This point is taken up later under 'Workload shifts'.

(iii) Some specific assessment methods are particularly good.

There were three specific assessment methods that were most regularly described as good. They were the MILE, the OSCEs and MEQs.

I like the MILE and it has been well-received by students. It convinces them that the skills developed in the PBL process are important skills; they realise that our assessment of these means that we take the skills seriously. [6]

The OSCEs really are a big improvement in the examination of clinical skills. And it's good they start as early as Year Two. OSCEs significantly reduce assessor bias in clinical assessment; their objectivity is a real plus. [7]

MEQs are difficult to produce, but they are a good method of assessment. They are clinically situated (which helps students see their point) and they can cover the whole curriculum, integrating different aspects, scientific understanding, medical and surgical knowledge, ethics... they really do work very well. [8]

Readers may wonder why EMIs and the MLC were not also singled out for praise. The answer is probably that these have not yet been tested, or adequately tested, in practice, given that the first students going through the new curriculum are currently in Year Five. These methods may well be singled out for approval in future, but at the time of writing people are reserving judgement and sometimes expressing reservations.

I am rather disappointed that there are no MCQs – but that may only be because I have written a book about them. They are now deemed to be politically incorrect, restrictive and punitive. I'm not sure about these EMSQs (EMIs) that are going to pop up in the finals for the first time; students will have had little or no experience of them. We'll just have to wait and see. [8]

The MLC looks good on paper. It should be a bit more standardisable and reliable than some earlier forms of clinical assessment. I'm not sure that one MLC will be enough; but it may be too time-consuming for examiners for us to have more than one. Some clinicians will feel the importance of their own specialism is not adequately recognised if all students are not thoroughly tested within it. [2]

(iv) There is a pleasing diversity and balance in the assessment methods deployed.

It is immediately obvious from Sections B(a) and B(b) that there is an impressive variety of assessment techniques deployed in this new curriculum. The diversity of assessment methods within the SSMs was especially favoured. There is a tendency in some assessment writing to assume that variety is a virtue in itself; this usually comes from those who see the over-dependence in some areas of higher education on essays and essay-type examinations as a reactionary anachronism. A majority of those interviewed liked the diversity in assessment in the new curriculum and defended it with rather more substantial arguments. Firstly the medical curriculum has a variety of aims (to develop understanding of the underpinning knowledge bases, to produce clinically competent graduates who are safe to practise, to encourage a commitment to independent learning and continuous professional development and so on); if assessment has to be sufficiently valid in relation to a range of purposes then assessment has to occur in diverse forms. Secondly, different methods of assessment are liked and disliked by different students: some produce their best work under examination pressure and conditions – others find these damagingly stressful; some students shine in practical assessment, others do not; oral presentations and examinations allow some to flourish and make others wither. There is therefore an argument from equity and fairness in support of diversity in assessment.

Some interviewees believed that the very diversity brought about balance within the assessment programme; others saw balance as requiring rather more than this. They approved of the relative weightings attached to different methods: 'no one method gets too much weight and unbalances the whole thing'. Furthermore, not only was the overall balance about right between assessing knowledge and understanding (an academic purpose) and assessing clinical capability (a professional purpose), but also the shift in emphasis from the former to the latter (while retaining a concern for both) was reflected in an appropriate shift in assessment methods.

(iv) It is good that there is more integrative assessment.

The assessment scheme was thought pleasing for its attention to integration. There were two main aspects of integration; the first was concerned with combining the academic and the professional (or the scientific and the clinical, or understanding with competence) and the second showed itself in the pulling together of contributory disciplines.

MEQs, as observed earlier, were recognised as effective integrating instruments, but most people tended to talk of integration as a pervasive principle of assessment planning rather than a characteristic of any one method.

Probably for the first time we are integrating different contributory disciplines. Whereas previously we had degree courses in X as a stand-alone subject we now have the different disciplines having questions in a joint exam paper, or better being integrated within questions. I think we are now actually assessing, throughout the curriculum, whether students can bring together all the different aspects of knowledge that they need to practise properly. ... [9]

(v) The assessment scheme as a whole is coherent.

Coherence was mentioned by five people; it should be noted that another four people expressed views about a perceived lack of coherence within the assessment programme.

People who liked the coherence of the assessment procedures attached two rather different meanings to the term. In the first of these, coherence was construed as closely related to validity and constructive alignment (of course purposes and assessment).

We can, I think, be justifiably proud of the coherence between what we tell students matters and what we then go on to assess. We tell them that we value X – and then they can see that we value X because we assess it. [1]

The second set of views was related to the existence of an underpinning conceptual organiser to which most of the individual assessments could be related – The Assessment Grid. This grid was not seen as perfect, indeed several people wanted it modified in different ways, but it was seen as providing a helpful, unifying framework which encouraged people, both staff and students, to see relationships between assessment elements that might otherwise appear somewhat disparate. It was pointed out that it was only if the scheme as a whole was coherent that it could become clear how it was also progressive i.e. how the demands of the course built on the expected achievements of earlier parts: ‘the whole thing has to hang together to help students develop’. Two people pointed out that although the grid had already shown its worth (perhaps most so during clinical attachments) there was still considerable scope for developing it further as a guide to student self-assessment and as a prompt to dialogue between tutors and students about strengths, weaknesses and progress.

Inevitably when talking about what was perceived as good, people sometimes identified what they saw as less-than-pleasing. There were five main things of this type; they can be summed up in these bald statements.

- Summative assessment is well-done; formative assessment is not.
- Students complain they don't know how well they are doing.
- The grading scheme should change.
- Students are under-assessed; staff are over-worked.
- There is a discontinuity in assessment at the end of Year Three.

C. TWO INTERPOLATIONS

(a) What recommendations did the GMC make about student assessment?

The new curriculum was planned to satisfy the policy recommendations of the 1993 Report from the General Medical Council, 'Tomorrow's Doctors'. What did that report have to say about how students should be assessed?

Only one of the principal recommendations mentions assessment. It states that 'Systems of assessment should be adapted to the new style curriculum, should encourage appropriate learning skills and should reduce emphasis on the uncritical acquisition of facts'. Of the sixty-eight paragraphs in the body of the report, only four are concerned with assessment. The main points in these paragraphs are as follows.

[para 53] The aim of correcting curriculum overload would be wholly frustrated if the present assessment systems were to continue. Students are reluctant to learn in areas in which they will not be assessed. The multiple choice format tends to emphasis acquisition of facts at the expense of reasoning. Assessment schemes must adequately test the achievement of the educational goals of the curriculum highlighted in the report and they must reflect the integrated nature of the curriculum.

[para 54] Assessment of the core curriculum must be rigorous to provide an assurance that the graduate is fit to take responsibility for the care of patients, albeit under supervision. A system of progressive assessment is required 'that monitors the acquisition and utilisation of core knowledge, that explores attitudes and that requires certification of the achievements of competence in the skills demanded by the course'.

[para 55] Assessment of the special study modules will require different, but no less important, procedures. Methods of assessment will vary according to the nature of the modules, but will often take the form of a short dissertation. Assessment of special study modules may provide the means of identifying outstanding achievement and may assist decisions on the award of honours and distinctions. The amount of work that will be required of assessors is acknowledged.

[para 56] Changes in assessment will require considerable modification of the existing roles of both internal and external examiners. Guidance, if not training, will be required for those who do the assessing under the new system.

(c) Clarification of terms

Some clarification of terms as used by the interviewees and as used in the rest of this report is necessary here. Firstly, some staff referred to continuous assessment and some to continual assessment. Although continual (in the sense of frequently recurring) is perhaps more accurate than continuous (connected throughout in time), conventional current usage favours continuous. In what follows the term continuous assessment is used. Secondly continuous is taken by the writer to be the opposite of terminal: continuous assessment takes place during a course (sometimes referred to earlier as 'coursework assessment'); terminal assessment occurs at the end of the course. These terms have to be very clearly distinguished from formative and summative. Thirdly, formative assessment is taken to be assessment which provides the student with some kind of feedback intended to influence their future performance; summative assessment is assessment which provides grades or marks which are taken into account in making end-of-course judgements of success and failure (or of levels of

success). Formative assessment is usually continuous, but continuous assessment may be summative as well as formative. Terminal assessment can have no formative impact during the course, but could well have a formative influence on subsequent, after-the-course learning. It was interesting to note that several interviewees used the term formative in a distinctive sense, one not found in other subject areas. They saw formative assessment as meaning 'practice assessment' i.e. students could be provided with a 'mock' examination which gave them a trial run for the real thing, but their result on that mock assessment would have no summative function.

D. STUDENT ASSESSMENT: LOOKING BACK

(a) Introduction

Interviewees were asked what they saw as the most significant changes that had occurred in assessment as the medical curriculum had moved from its previous form to its new one. They were then asked why they thought such change had taken place. It was anticipated that this would be a difficult question to answer and so it proved. Change usually has a multiplicity of causes and these can be subtle and interacting. It is possible to be involved in change and not to know, or much care, why it is happening; it is also possible that one knew at the time but cannot now recall the reasons, or indeed that one's past perception of what the powerful factors were has been modified with time. However, no apology is necessary for asking the question; people's perceptions of reasons are arguably as important as the 'real' reasons, even if it could be agreed that these are ever accurately knowable. There was an additional complication in this subject area of medicine. There was one general answer available ('Assessment changed because the course as a whole was re-designed') which meant that people tended not to go on to deal with why it was that particular changes had occurred.

In the account which follows, there is some repetition of points made in Section B; this is because what people saw as pleasing features of assessment in the new curriculum tended to influence what they then identified as significant changes from the previous curriculum

(b) Significant changes in assessment

Several people prefaced their answers by making a distinction between the big changes in assessment that were part of the original curriculum planning and the later modifications that were made as the assessment system was implemented. The latter were still continuing. The difference between 'major original' and 'minor subsequent' was not quite the same as 'more significant' and 'less significant'; cumulative minor modifications might in the long-term prove to be highly significant.

There is no comparison between assessment in the old and in the new. But, I think I want to distinguish between the original plans and further evolutionary change. The assessment scheme is certainly not static. It was conceived before any actual students appeared and assessment practice has of course been modified. [6]

There were four changes described as 'most significant' by a majority of those interviewed: less assessment of factual knowledge; more integrated assessment; more standardised clinical assessment and reduced dependence on 'conventional' assessment methods.

Less assessment of factual knowledge

It was regularly said that the former curriculum had been over-loaded with content and had over-used content assessment. Traditionally, medical students were 'tested on a vast bundle of knowledge' and then they appeared to use only a small fraction of that bundle in their later practice of medicine. Research had indicated that the rate of forgetting after final examinations was astonishingly high. However, some interviewees stressed that they were not just talking about a reduction in the size of the knowledge base that was assessed; there had been a very serious attempt to identify the knowledge that would be **useful** to students. The aim had been not merely to reduce the burden of content knowledge assessment, but to prune it judiciously so that students would be more confident that they possessed the understanding necessary to tackle the problems that patients would most commonly present them with.

Other interviewees emphasised that a reduction in the assessment of factual knowledge meant an increase in the assessment of higher cognitive abilities i.e. a shift in balance. The higher ability most often mentioned was 'the ability to provide an informed critique'; such an ability showed itself in the capacity to analyse clinically situated scenarios and problems and then go on to weigh up the strengths and weaknesses of different theories, diagnoses and patient treatment/management strategies. The changes in assessment that people described can be interpreted as 'Bloom' category shifts: from categories 1 and 2 (Knowledge and Comprehension) either to category 3 (Application) or to categories 4 and 6 (Analysis and Evaluation). This should also be seen in relation to the GMC's principal recommendation on assessment (see section C a): 'Systems of assessment should reduce emphasis on the uncritical acquisition of facts'.

There was a general assumption that this significant change was 'a good thing'. One person, however, disagreed. The change was not good, because of the number of students who, s/he said, complained about it: 'We want much more assessment like in the old curriculum'. The student perspective on assessment is not a part of this research; there is an obvious task here for other researchers. It should be noted, however, that there was an ambiguity in this one person's views. It was unclear whether students were saying they wanted **more** assessment or they wanted more **feedback** from assessment on their individual performance level and their relative standing amongst their peers.

More integrated assessment

In the past it had seemed as if there was a separate assessment episode for every subject area and topic within the curriculum. This had changed.

We no longer have a departmentally-based course and we no longer have subject-based assessments. There has been a big shift towards assessment which is multi-disciplinary and integrated. [8]

This category of responses claims to be about 'integrated' assessment. However, it is possible to detect distinctions and differences of emphasis across the individual responses. Some people were referring to what could be called **integrative** assessment, where one assessment method required students to pull-together understandings from different subject areas, disciplines, or specialisms. A good example of this was the MEQ which began with one patient-based scenario and then went on to sets of questions carefully structured to draw on student understanding of a range of underpinning knowledge bases. The assessment method could reasonably be claimed to avoid compartmentalisation of knowledge and to be integrative. Other people described what we may call **integrable** assessments; they perceived a requirement within the new curriculum to relate individual assessment instruments, methods

and episodes to some organising and unifying principle or schema. One pervasive principle was that assessment should reflect the reality of effective clinical performance. This required not merely the deployment of 'deep' knowledge of one highly specialised area, but also the wise and creative drawing together of knowledge from a wide range of relevant contributing disciplines, scientific, social and ethical. An organising schema for assessment was provided by The Assessment Grid. When designing any separate bit of assessment, the designer had to bear in mind the assessment domains and levels made explicit in the Grid. People, it was said, had become more aware of the need to see the parts of assessment for which they were responsible as part of some greater whole. Overall coherence of the assessment scheme was more likely when individual assessments were integrable.

It was noted more than once that the change to more integrated assessment had resulted in significant changes in who actually did the assessment i.e. who designed it, who managed it and who marked it. This is discussed further in Section G (b).

More standardised clinical assessment

Clinical assessment within the previous curriculum was described as having been, on the whole, unstructured and chancy: 'It was pretty much the luck of the draw what patients and medical conditions the students met and what examiners they encountered'.

In the old long case assessment, students spent an hour with the patient unobserved and then talked with the examiners for 20 minutes. This could be a good exercise when it was well-examined; it depended very heavily on the skills of individual examiners and it could on occasion descend into a trivial exercise. Some patients were more co-operative than others and the examiners didn't know what the patients had been like. And obviously some diseases are more complex than others. In the past, the short cases when they were well-done were very searching. But there were problems; not only did the patients vary a lot but there was a lot of variability across the large groups of examiners that had to be used. [9]

Within the new curriculum, the use of Objective Structured Clinical Examinations from an early stage and the introduction of the Modified Long Case to Year Five had together brought about a very significant change in clinical assessment. These methods were seen as more controlled, more standardised and more reproducible and allowing more consistent professional judgement; in addition, and perhaps just as importantly, their introduction and design had required a serious exercise in the clarification and explication of the criteria against which student performance would be assessed. In the conventional language of assessment these methods attempted to increase the reliability of assessment, while maintaining authenticity and validity. Considerations of equity and fairness to students were given as important factors influencing the changes.

Reduced dependence on 'traditional' assessment methods.

In the previous curriculum there had been over-reliance on two assessment methods: long essay-type questions and multiple-choice questions. These had been replaced by modified essay questions (and short notes) and extended matching items. These changes were generally approved. MEQs avoided rambling, discursive bits of writing on broad topics and themes and encouraged integration of knowledge from different sources; short notes emphasised the virtues of economy and succinctness; EMIs were likely to require of students more than the recall of specific bits of knowledge. Implied in everyone's responses, and made wholly explicit in two cases, was a recognition that methods of assessment give strong messages to students about the type of learning that is required of them; there was a good chance that

assessment methods in the new curriculum would have a positive 'backwash' effect on the quality of student learning.

I think that the methods of assessment we have now adopted will actually help students to learn more effectively. I have no doubt that essays and MCQs encourage superficial and forgettable learning and a surface approach to learning. Students are beginning to get the message that low-level factual knowledge is just not enough.

[1]

There were two other changes seen as significant, not by a majority but by three people each. The first, although variously expressed, can be described as increased alignment. These interviewees said that it was much more obvious in the new curriculum that assessment methods had been brought into line with the stated purposes of the curriculum. Official documentation described the main aims and the intended learning outcomes and it was obvious to both medical teachers and to students that what was officially said to be important was then given important weighting in the assessment procedures. This is what was described earlier (page 4) as a close correspondence between the rhetoric of course purposes and the reality of assessment practice. Just one example of this was the MILE; independent learning was said to be important and its importance was then confirmed by the existence of an assessment of independent learning. The second change was said to have been a shift to more continuous summative assessment; the existence of 'necessary hurdle' assessments and coursework assessments had both helped, albeit to a rather limited extent, to reduce the stressfulness of crucial big-bang, terminal examinations.

One further change was identified, this by one person only. S/he thought it highly significant that the principle of formative assessment had been clearly acknowledged, even although its practice was distinctly patchy.

(c) The reasons for these changes.

At the outset of this research there was a hope that people would identify the change that they saw as most significant and then reveal what they saw as bringing about that change. It may be that this hope was naïve or misconceived, given the complex nature of causality in human affairs, or that the interviewing technique was inadequate to the task. What happened in the first two subject areas explored was that two or three significant changes were described and then people identified what they saw as important general factors that had influenced the changes collectively. It was not normally possible to estimate the potency of any one factor in relation to one specific change. In medicine the situation was somewhat different. As indicated earlier (top of page 16) those interviewed in medicine had one very obvious answer to hand and did not usually see any need to go beyond it. Given the amount of ground to be covered in a limited time, the interviewer did not press people at this juncture. However, more was revealed about the factors influencing change later in the interview; this is reported in Section F.

Why was assessment in the new curriculum significantly different? 'Because the whole curriculum had been re-designed and inevitably assessment was part of that process...' The answer was so clear that the question began to seem not worth asking. But why was it these particular changes that had been introduced? Not everyone expressed a view, but the reasons that were given were of four types: general dissatisfaction with the past; responding to an external (GMC) policy 'steer'; giving students the right message about what learning was required; greater harmony with the nature of medical practice. Dissatisfaction with past assessment procedures focused both on the quantity of assessment ('students were horrendously over-assessed') and on the over-use of flawed assessment methods ('unfocused

essays', 'trivial MCQs' and 'non-standard clinical cases'). The 1993 GMC report was recalled as having directed (a) that there should be less assessment and (b) that there should be less emphasis on assessing whether students possessed knowledge and more on whether they could apply it sensibly. Assessment was recognised as sending powerful messages to students about the type and quality of learning expected of them; given the right kind of assessment students should become better at deploying useful (and less forgettable) knowledge bases flexibly.

Several of those interviewed said that assessment had changed because of the need to bring assessment methods into closer harmony with what students would be doing after they graduated/qualified: 'How students are assessed should be strongly influenced by what they will be doing as pre-registration house officers'. There is, however, a distinction that ought to be made between aligning assessment better with the nature of future employment and aligning it better with educational purposes. One interviewee made this very clear.

Yes, assessment does have to be informed by what we see as the current practice of medicine. But we do want to keep pressing forward. You don't want what doctors actually do to remain the same. We should be trying not to look at the existing state of medicine – but at some future more desirable state. We have to broaden students' horizons, to educate them beyond utilitarian needs. Competent doctors are something rather less than professionals who have been educated for future demands. [3]

This quotation raises profound questions about the purposes of higher education and the nature of professionalism. Within this new curriculum are students being assessed on their fitness to practise or on the extent to which they have benefited from higher education? Should professionalism be so conceptualised as to require its entrants to be both proficient within the existing paradigm and critical of it?

E. STUDENT ASSESSMENT; LOOKING FORWARD

(a) Changes wished

Interviewees were asked what changes they would personally like to see in how students were assessed, assuming that the conditions for introducing such changes could be made favourable. Those interviewed had a lot to say; responses were thoughtful and rich in ideas. The general view (with perhaps one exception) was, 'It's pretty good, all things considered, but it could be even better'. There was no sense of people wanting a lot of change because assessment was in a bad way. It was also striking that people were not saying 'Some kind of change is needed; they should do something about it', but rather 'I know what kind of change is needed and I'll be doing something about it'. It would be foolish to exaggerate this positive feeling that people seemed to have of being personally able 'to make a difference' (and the evidence is of course only coming from ten or eleven people), but this researcher was forcibly struck by the continuing enthusiasm for assessment improvement of hard-worked people whose practice has already gone through considerable upheaval. This is in sharp distinction to some corners of the University in which staff appeared not to have sufficient interest in assessment even to be described as apathetic. The Faculty of Medicine deserves praise for managing to retain this enthusiasm through a very demanding period and it should be seeking to harness the energy and goodwill that is around, further to develop and enhance assessment practice.

Some people focused on one big innovation they wished to see ('We need a Maastricht-type Progress Test'); others listed half-a-dozen 'tweakings' of existing practice. However almost all the recommendations for change can be related to three broad themes:

giving students a better idea of where they are;
better review of coherence of assessment;
further improvements in clinical assessment.

In addition four interviewees emphasised changes they wished to see in Year Four assessment; this will be explored briefly as one specific focus in which all three themes can be discerned. All the other suggestions, each of which came from at most two people, are listed in one final paragraph.

(i) Giving students a better idea where they are.

Almost all of those interviewed spoke of assessment changes they wanted that could be related to this general theme. However, there was no simple consensus about what this meant or how to do it.

People frequently mentioned formative assessment: 'I want a change to more, and better, formative assessment'; 'All the hard thinking went into summative assessment; formative assessment is very patchy indeed'. About half of these people construed formative assessment as outlined in Section C: any assessment was formative if one important purpose was to give students useful feedback about their performance. The other half attached a distinctive meaning to the term; they saw it only in terms of 'practice assessment', provided in the same form students would later encounter as summative assessment – first the 'mock' and then the 'real'. There are rather different underlying assumptions here; in the first case the benefit of formative assessment is assumed to lie in the helpful effect on future learning of feedback provided by staff; in the latter case students are assumed to benefit from what they learn for themselves from practising a form of assessment, i.e. having a dry-run.

Formative assessment. For example, we could do diagnostic computer-based assessment of the basics, to allow later summative assessments to get at the higher order things. It would be really helpful for students to have quick feedback that they at least knew the basic stuff. [10]

Whatever we do in the finals, I want it to be done in a formative way first – in a mock exam. [4]

Staff also gave different reasons for wanting more (and better) formative assessment: 'All students need to know that they are learning the right sort of stuff to the right level'; 'Medical students are highly competitive – they need to know how they are getting on compared with others'; 'Students need to know they are making progress'. These views are particularly interesting in that they suggest three different ways of interpreting the title theme. Students should have a better idea of where they are. Yes – but in relation to what? To a defined set of learning outcomes or a knowledge domain? To their peers i.e. in relation to fellow students in their year? Or to themselves at an earlier stage? There is a pleasing parallel here with three fundamental approaches to assessment, the criterion-referenced, the norm-referenced and the ipsative.

One person articulated a somewhat different reason for more formative assessment, that it would decrease the worrying stress levels amongst medical students.

There is research being done in Glasgow about the stress students experience in coping with the new curriculum. There is no doubt that assessment is an important stressor. Students are significantly more stressed as summative end-of-year examinations approach. Between a quarter and a half of students are stressed at the end of the year. We have cut down on the overall assessment, so students now feel more stressed about the fewer summative assessments. The answer is not to reintroduce more summative assessments – but to have better formative assessment. Not knowing how you are doing, how you are going to do, is stressful.... [3]

Also related to this theme were several recommendations about desired changes in how the results of assessment are reported to students. The key point made here was that when an assessment result was communicated to students as a simple binary decision of 'satisfactory' or 'not yet satisfactory', then staff were routinely met by requests for 'a better idea of how I'm doing'. 'Was I only just satisfactory or securely satisfactory?' 'Was I so exceptionally satisfactory that you really ought to say I'm distinguished?' 'Was I nearly satisfactory or lamentably unsatisfactory?'

Staff wanted assessment changes that would give students a 'better idea of where they are'. Given the range of meanings involved and the variety of reasons adduced for the change, it is unsurprising that there were a large number of specific recommendations made for change. These included the following: encourage staff to spend more time to provide feedback; train them to give more useful feedback; design high-tech feedback systems using computer comment banks; devise helpful interim assessments without a summative role; provide more 'mock' examinations; change the grading scheme; improve the recording and reporting procedures. There was one change, mentioned by three people and described at length by one: the introduction of a Maastricht-type Progress Test. This may be unfamiliar to some readers and it seems to deserve being dealt with at some length.

The Progress Test as developed at Maastricht¹ depends on a huge bank of questions, with associated performance data, from which can be selected a battery of items which cover the whole field of medicine. It is relatively easy to construct a series of tests conforming to a pre-defined test specification. Such tests are given to all students, four times a year, throughout their course beginning about three months into the first year of study. On the first test they sit, students may score, for example, 10%, but by the end of their course average scores are around 90%. Although the assessment orthodoxy is that you should not give people tests on which they can only achieve very low scores, students accept that low scores are only to be expected in the early stages and do not find them de-motivating; later confirmation that they are progressing is highly encouraging. Because of the size of the item bank, questions cannot be spotted and the test can not be specifically prepared for. After each sitting of the test, each student gets a computer generated profile with comments on their progress. Students whose progress appears less than satisfactory are required to talk with tutors. It is claimed that student attitudes to the Progress Test are very positive and that it has two very significant advantages: (a) all students are informed regularly, and from an early stage, 'where they are' in relation to the medical knowledge field, to their peers and to their own earlier position and (b) it allows the complete abolition of big-bang terminal written examinations (but not OSCEs and other clinical assessments).

Proposals have been made for a Glasgow Progress Test. It is argued that Maastricht does not use the most effective question format ('it uses bog-standard MCQs'). The challenge is thus to create a large item bank, perhaps of EMIs, and perhaps via co-operation with other medical schools in the United Kingdom. The assessments could be computer-mediated. The

¹ See, for example, the Assessment Programme as described in *The New Maastricht Curriculum* (2001) Institute of Medical Education, Universiteit Maastricht.

performance data generated would not only serve the immediate interests of individual students, but they would also help answer a range of course evaluation and accountability questions and interesting research questions, relating for example to performance levels over time and to inter-institutional and international comparisons.

(ii) Better review of coherence of assessment

Desired changes here were described by rather fewer people, but still by a majority. People went into less detail on this topic than on the previous one, but seemed to feel just as strongly about it.

We need an assessment review group. There was one – but now that things are up and running, people have not found the time to keep it working. The curriculum and assessment have developed year by year; Years One and Two were happening when Years Four and Five were only on paper. There needs to be a closer look at the whole programme of assessments. Have we got coherence across the whole scheme? Very soon we must look at this. I have ideas; block leaders have ideas, year co-ordinators have ideas. Whether what we are doing ties in well with what's going on in the other years is a bit unknown. Are students getting consistent messages across the years?

[6]

We need an assessment group that deals with everything to do with assessment across the whole curriculum. There was one, in the early stages; it got bogged down in political correctness and arguments about different techniques. It stopped; the chair resigned and it never re-established itself. So the problem as I see it is that we are supposed to have an integrated curriculum from Years One to Five. Years One to Three are quite similar; there's nothing much in Year Four except clinical appraisals; then the finals introduce a totally new type of question format – students have no experience of it and it doesn't reflect how they have been taught. I expect our new Associate Dean will get to grips with this after the first set of finals this June....

[8]

Although the most frequently expressed arguments for a review group were related to coherence and progression within the assessment scheme, such a group was also thought necessary by some interviewees to serve other purposes; encouraging research in assessment, initiating developments in assessment and making more use of student assessment in course evaluation.

I still think there needs to be some body looking at assessment overall. It is understood extremely well in the context of medical research that evaluating an innovation in treatment is meaningless without the right outcome measures. There is not the same realisation of the importance of assessment. Perhaps some of my colleagues realise this, but they are so busy with clinical commitments and their own medical research that education is not as high on their agendas as it might be. It is easier to deliver education than it is to assess students and to evaluate programmes. That's why we need a review group....

[3]

(iii) Further improvements in clinical assessment

A majority of those interviewed identified some change they would like to see in clinical assessment; suggestions on this theme were made much more tentatively than on the previous one, probably due to the as yet not-much-tested nature of the clinical assessment

arrangements. There was a very reasonable desire to see how the assessment arrangements 'bedded-down' with the first set of Year Five students before coming out with strongly worded proposals for change.

Changes that people wished were of three types. Firstly, there should be more sampling of clinical and practical skills; too small a sample of these were currently included in OSCEs. Assessment should require the demonstration of clinical competence across a wider range of skills. Secondly, and probably related, the range of specialisms represented in the MLCs should be reviewed and then extended. Thirdly, more use of 'standardised' patients was required; it should be possible to 'train' patients better to present a similar level of challenge to a number of students. It was very clear that everyone wished change in clinical assessment to satisfy three criteria (a) there should be no reduction in the authenticity of such assessment, (b) there should be an increase in the reliability of such assessment and (c) there should be no increase in the assessment burden falling on clinicians. Whether or not an enhanced assessment scheme could ever be devised which would satisfy all three criteria is arguable. 'You want assessment that is authentic, reliable and cheap? You can't have it. You have to settle for any two out of the three.'

What of Year Four?

Four of the eleven interviewed specified some change they wished to see in Year Four assessment. Although the division between the pre-clinical and clinical years is less clear than it once was, there was still a perception that Years Four and Five made up one two-year clinically-based period that was distinctively different in character from Years One to Three. A great deal of attention had been directed at what happens at the end of Year Five ('the finals'); it was thus unsurprising that some people said that end-of-year Four assessment deserves to be the focus of more thought. There was some unease that students might not be sufficiently rigorously assessed and might not know early enough if their clinical progress was unsatisfactory.

In Years Four and Five end-of-block assessment is not that rigorous. We probably need more rigour in Year Four and that may mean introducing some more formal assessment at the end of Year Four. Year Four students could be allowed to sit the Year Five examinations for formative purposes. We could introduce some modified long cases at the end of Year Four blocks. We have to ask if we need summative assessment at the end of Year Four. In my view the resources for this are just not available. [9]

I know that the two years of clinical training have to be seen as a continuum, but I do think we need more formal assessment at the end of Year Four. Without that you get students with problems in Year Five and that is really much too late in the day to tell them 'You are not going to make the grade'. I would extend this comment: I think we need earlier and better identification at every stage of those students of whom it can be said 'This is not the course for you'. [7]

There were four desired changes in assessment, each mentioned by one or two people only. These were the following: more weight to be attached to portfolio assessment in Years Four and Five; the introduction of computer-based diagnostic testing; the re-introduction of exemptions from written examinations and more serious attempts to tackle the problems of assessing ethical dimensions.

(b) Changes that may be required.

The previous section dealt with answers to the question of what assessment changes people would personally like to see; in this section the prompting question was 'What changes in assessment do you think you may be required to make, whether you like them or not?' Most people started off by saying 'Nothing very much' or 'There's nothing in the offing that I know of' and then went on to talk at some length. All the changes that people foresaw involved extra-faculty pressure and policy, except one.

I think the faculty is about to embark on a monitoring exercise; we'll examine all the existing statistics for student performance. These may point to some changes that are necessary. There may be a pressure to change from our internal monitoring activity. [1]

The phrase 'extra-faculty pressure and policy' has just been used; it deserves some unpacking. From where, outside the faculty of medicine, could pressure and policy come? Is there some useful distinction between pressure and policy? The first question can be answered at the outset from the responses of interviewees. An attempt to answer the second will be delayed to the end of this section; in the interim the term 'requirement' will be used somewhat loosely.

Those interviewed four extra-faculty sources that might possibly require changes in assessment in the future: The university (nine mentions), the General Medical Council (six mentions), the Government (two mentions) and the Quality Assurance Agency (one mention). The word source may itself be misleading; it was noted by some that the apparent source might well not be the original one. For example, a policy which originated with Government might well be mediated through the General Medical Council; a policy which appeared to come from an institutional source (The University) might have been required of that institution by the QAA.

It was well-known that within The University there was a Working Group on assessment, reporting through the Education Committee to Senate. Nine of the eleven interviewed thought it possible that The University might require the Faculty of Medicine to use a grading scale that would be uniformly adopted across The University. Reactions to this possibility covered a whole spectrum of opinion.

Imminent is the development of a university-wide Code of Assessment, the resistance that exists within this faculty to using a grading scale may well be over-ruled by that. There may be tensions if everyone has to use a 20-point grading scale, but from my perspective such usage will be helpful. [7]

The University pressure for standardised grading might force a change to grading clinical performance that seems to me inappropriate, although it may be quite reasonable for written academic work. I think the convener of the Working Group understands our position on this, so university pressure worries me slightly but not excessively. [2]

Students have to achieve understanding of the defined core of medical knowledge; either they do or they don't. Whether or not they do is a matter of informed professional judgement. We have successfully resisted banding, so we will resist the spreading of students out across a twenty point scale. [10]

Any proposed grading scheme which suggests that academic judgements about the level of achievement of intended learning outcomes must be translated into a series of numbers is a nonsense and must be very strongly resisted. [1]

It may at first be thought that something as apparently minor as a grading scale can arouse such strong and diverse reactions. Not so. People held strongly to two distinct philosophies, or ideologies, about assessment. The Working Group was perceived as requiring, via a common grading scale, a way of relating two assessment paradigms which may at base be incompatible. Either this was viewed as a flawed but necessary compromise that could be made to work, or as an unholy union of opposites that could only be approved if rationality was suspended. There were also, however, hints in the reactions that the objection was not so much to **this** grading scale, but to the imposition in the name of uniformity and consistency of **any** grading scale; after all the faculty of medicine is distinctively different from others, it has achieved a highly-acclaimed revolution on curriculum and assessment and it is a high status faculty.

I really don't see why they want uniformity across faculties. There is only a strong argument for comparability within faculties. [11]

This whole debate is dealt with at greater length as 'The 2,3,4,7,20,100' issue in Section G (a).

Six of the eleven interviewed said that future change in assessment might be required by the General Medical Council. The whole reason for the new curriculum and the new assessment scheme lies in the GMC document of 1993. The GMC had been saying similar things for the previous twenty years and 'no-one had paid a blind bit of notice'; thus the 1993 document said things more forcefully than hitherto. If there was evidence that medical schools had not moved sufficiently since 1993, then it was possible that the next major policy document from the GMC would be more forceful, directive and prescriptive and that its policy imperatives would deal specifically with assessment.

One person illustrated this point with regard to oral examinations.

The GMC makes what are essentially recommendations; it has not in the past been directly prescriptive with regard to assessment. I think it said in 1993 things like 'we are not over-enthusiastic about oral exams', but that's not a direct prohibition telling you that you must not have them or a prescription about how they must be done. I wouldn't be surprised if they exert more leverage in future. [3]

The GMC was seen as the mediator of government-derived policy and the interpreter of public opinion. The public had rising expectations of what constituted the competent doctor and the GMC would not, for example, wish any doctor to be perceived as incompetent because they were merely incompetent communicators. This might then be translated into a requirement for change in assessment priorities and methods.

Each year there was a very small number of students who were academically and technically sound, but were unsuited to the practice of medicine because of their attitude and approach. It might be necessary to use Practice Panels which effectively de-couple graduation from registration. This could have profound repercussions on the assessment of medical students.

Other people mentioned rather less momentous matters and ones that did not suggest any great alteration of the role of the GMC in relation to assessment change.

The GMC were happy with us last time. But will they still be happy with our assessment of clinical skills? [8]

Everything the GMC encouraged us to do, we have done. Even if not well enough. I would not be surprised if they come up with recommendations for best practice in assessment. We would take account of those, but recommendations are hardly requirements. [1]

The government was mentioned briefly and only twice. The Government might require two developments in the interests of 'efficiency savings', (i) a mechanism for allowing the best students to 'fast-track' the curriculum and finish earlier and (ii) co-operation with other medical centres in the production of computer-mediated and other assessment materials. Both developments would have significant impact on how assessment was carried out.

The QAA was quoted by one person as likely to require some change. This was not, as might have been predicted, because of some precept in the assessment section of their code of practice.

Teaching Quality Assessment or Academic Review or whatever it's called nowadays is coming round again. I'm not as confident as I was last time because we don't have an overarching assessment committee and we don't have good enough documentation showing how it all fits together. [8]

Is pressure then to be distinguished from policy? Those interviewed seemed to talk of pressure as anything in the environment that made it likely that a change of some kind would be made. Pressure might well derive from policy statements, but it could come from other sources, for example, sensitivity to staff, student and public opinion.. Policy was seen as a more tangible, formal thing which derived from some body which had some legitimate authority over one. Policy might prescribe a particular set of actions and it might proscribe others. (For example, in future The University might require the use of one grading scale and forbid another.) On the other hand policy might only require that a general precept or principle be accepted and that evidence be made available to support a claim that it had been accepted in some form or another. (As in the QAA Code.) The details of what was adopted might vary from context to context. Policy then was not just to be seen as a directive from on high that had to be implemented in detail; policy might well allow scope for interpretation and re-construction. People also distinguished between policy production and policy enactment; the former was policy activity in 'a contested terrain' in which both pro-activity and resistance were appropriate. And policy which was enacted should not be seen as set in stone for all time...

F. ASSESSMENT: FACTORS INFLUENCING CHANGE

In Section D (c) there was some consideration of the factors that had influenced the changes in assessment introduced into the new curriculum; in Section E (b) there was an indication of factors that people thought might become more influential in future.

After trying to elicit people's own views on factors influencing assessment, the interviewer presented each person with a list of factors identified by others and asked for responses to these. Interviewees were asked (a) to give their view of the importance of these factors in relation to assessment change in medicine in the past and (b) to predict which would become more influential in future.

It was very noticeable that everyone provided crisp and decisive responses. On only four factors was there any extended comment. There was almost complete unanimity that four factors had been very important and that eight had been of no importance. On only three factors was there any spread of opinion. The results are summarised below.

Factors influencing past change in assessment

Factors for which the most frequent response was 'very important in medicine'

- A change in course structure
- Policy from an external professional body (the GMC)
- A general feeling that change was desirable
- Concern for the quality of student learning

Factors for which the most frequent response was 'of some importance' (and on which there was a spread of opinion)

- One enthusiast persuaded us
- Awareness of trends in assessment elsewhere
- Educational/assessment theory or theorist

Factors for which the most frequent response was 'not important in medicine'

- Students pressed for change
- Institutional policy required change
- National policy required change
- The availability of appropriate assessment technology
- Emphasis on transferable skills
- Pressure from external examiners
- Increased number of students
- Increased diversity within student population

Factors predicted to become more influential in future

- Institutional policy (The University)
- Policy from an external professional body (The GMC)
- National policy (The Government and QAA)
- Student pressure

N.B. Several people added that one factor would now **decline** in importance – the general feeling that change was necessary.

Only three factors provoked both spread of response and comment. The suggestion that 'one enthusiast persuaded us' elicited several remarks that there had been not one enthusiast but several. The word enthusiast itself made two people uneasy: it suggested both a 'champion' and 'a nutter with a passion'; there had been the former but (probably) not the latter. It was also thought that the enthusiasts(s) had not created awareness of the need for change but had influenced the nature of the particular changes it was decided to introduce. Secondly there was some ambivalence in people's reactions to 'awareness of trends in assessment elsewhere'. It was acknowledged that developments in medical schools elsewhere (especially Maastricht, Canada and New South Wales) had alerted people to the kinds of change which were possible, but people stressed that Glasgow (while avoiding parochialism) had not simply

'fallen into line' with things done anywhere else. Thirdly, there was a range of opinion on the influence of educational theory and theorists; this was distinctly different from the two previous subject areas where it was almost unanimous that educational theory and theorists were of no significance whatsoever. As with the responses to the 'enthusiast' factor, people thought that theory had not precipitated change, but had significantly affected the form it had taken. It is probably significant that there is a substantial body of research in medical education and the assessment of medical students; it was the literature from this that a few people quoted as influential and not the general education/assessment literature. This writer found it very intriguing that these people seemed to be saying not only that theory mattered, but that they themselves had no objection to being perceived as theorists. It may be of course that the interviewer was perceived as some kind of educational theorist, someone who would be pleased to find acknowledgement of the place of theory. Perhaps – but this certainly did not happen in any of the other subject areas explored.

The single factor on which there was some comment, but no spread of opinion as to past significance, was the availability of appropriate assessment technology. Computer technology was not seen as having been important in influencing changes in how students were assessed. Two people commented that there had been considerable growth in the use of computer analysis of assessment statistics and in recording and reporting the results of assessments.; another two noted that there had been computer-marking in the past (of MCQs) but this had now stopped; another said there was no place for any computer-mediated assessment within medicine at present, but the rapid developments taking place in this area might well make it appropriate in the future (perhaps in a Glasgow Progress Test). Although the researcher had been thinking primarily of computer technology, two people said that there was technology used in various forms within OSCEs, for example mannequins were used to simulate parts of the human body.

Readers may be surprised at the long list of factors seen as unimportant. The number of medical students has not changed in the recent past; number of students had no influence on assessment change. Admission to medicine is highly selective, thus the student body was not seen as more diverse in its abilities. Although there are many different career paths within medicine, the medical curriculum was not seen as preparing students for any employment other than medicine; thus pressure to give more attention to the assessment of transferable skills was not a factor in change. (There had been much increased emphasis on communication skills and their assessment, but these were increasingly seen as key skills within medicine rather than skills that can be transferred to other jobs.)

All but one of the factors seen as likely to increase in importance were related to increased policy demands from external sources, where increased meant both more of them, and each of a more directive nature. The fourth was student pressure. Partly this was the general view, expressed in other subject areas, that students were everywhere becoming more conscious of their rights as customers and more frequently litigious. (There were, for example, more appeals against assessment decisions within the new curriculum.) It was also, however, seen as much more than this. The curriculum as a whole stressed the production of independent-minded autonomous learners; the curriculum would not be genuine in its student-centredness unless it attended more carefully than it had in the past to 'the student voice'.

G. EMERGING ISSUES

In this section four issues are identified and expressed as questions.

- (a) What is the appropriate grading scale?
- (b) What do workload shifts mean for the management of assessment?
- (c) What kind of expertise do assessors need?
- (d) Can assessment be both authentic and reliable?
- (e)

The researcher had no pre-conceived idea that these would emerge as important issues; questions about them were not included in the interview schedule. These 'emerging issues' are subjects of debate seen as important by the people interviewed.

(a) What is the appropriate grading scale to use? The 2, 3, 4, 7, 20, 100 issue.

This issue has already been introduced under the heading of a change that may in future be required by The University. (Section E(b)) From the interview data it was not merely emergent or salient but positively irruptive.

It is difficult to get to the nub of this. Over time the issue appears to have changed from a matter that no-one thought much about, to one in which an internal orthodoxy was temporarily accepted, to one in which some people changed their views in the light of experience and then to one in which the possibility of external intervention prompted emotional attachment to positions not easily reconciled. At the time of writing it is impossible to see how the question may be answered and the issue resolved.

An attempt is being made here to keep the issue intelligible by referring to the use of a grading scale. It should be borne in mind that this conflates, and perhaps confuses, at least six separate processes in which some grading scale has a role to play: putting a summary grade on a piece of assessed work or an episode of assessed performance; calculating an aggregated grade for a collection of assessed elements; conveying to students their absolute or relative level of performance; generating an assessment profile to encourage self-assessment and to relate separate assessment one to another; conveying to a central registry summary grades on each reportable component of the course; publishing a final verdict for graduation and qualification. Perhaps an illustration, although in some aspects fictitious and contrived, may help to clarify these processes and also the numbers in the sub-heading.

- (a) A student is given a mark out of 20 for one part of an MEQ.
- (b) The MEQ has five parts and the student accumulates a mark of 90%. This level of performance is judged as Excellent and the student is awarded a grade A (on a seven band University grading scale of A to G).
- (c) The student seeks to know better how s/he stands. The tutor reveals 'not only were you an A you scored 90% and were the best student in the year'.
- (d) At the end of a clinical period, a student's preliminary self-assessment against seven criteria at three grades (Unsatisfactory/Satisfactory/Highly Satisfactory) is confirmed or otherwise by the educational supervisor's single summary judgement: 'Satisfactory'. The student presses the supervisor for more and is told 'You are satisfactory, but only just. Personally I think there should be four grades not three.'
- (e) Note to registry: 'In this component of the course we make a simple binary judgement, yes or no. This student has satisfied the requirements – Pass.'
- (f) Letter to graduand: 'You are judged by the Board of Examiners to be fit to practise. However, because of your outstanding performance in some areas, you will be awarded the degree with Commendation.'

There seems little doubt that the original planners were convinced that assessment judgements should whenever possible be simple binary ones: in the professional judgement of assessors either the student reached the required standard or s/he did not. There was an underlying assumption that assessment would be carried out against clear and explicit statements of intended learning outcomes. The appropriate grades were thus Pass/Fail, Satisfactory/Unsatisfactory or Competent/ Not-Yet-Competent. The underpinning philosophy, or ideology, of assessment was criterion-referenced and competence-based. This was clearly appropriate for clinical capability; at the end of the period of 'training' either the student was fit and safe to practise or s/he was not. This way of thinking was also extended to more academic concerns. There was a defined core of necessary knowledge that students must possess; either they did or they did not. It might be that in some single assessment a student would emerge with a mark, with a number attached to their performance, but what did a mark of, say, 65 mean? It remained a matter of professional judgement as to whether this mark indicated a satisfactory level of knowledge/understanding or not.

At some stage this attachment to binary judgements was weakened. This seemed to have come about for at least four reasons. Firstly, students were dissatisfied with a minimal response to their assessed efforts: 'Students want much more information on where they stand'. Secondly, it was felt that rigid adherence to this model would disadvantage Glasgow students when they competed later with the products of, say, Edinburgh: 'It would be grossly unfair if a Glasgow graduate could only say "I'm satisfactory" when the Edinburgh one could say "I have this medal" '. Thirdly, it was argued that the selectors of students to pursue intercalated degrees would 'demand to know which of the students were the high-flyers'. The fourth argument was more theoretical. In the assessment of academic work judgements could be made about levels of understanding and about the security of achievement of higher cognitive abilities; it was thus appropriate to have a range of grades available. Clinical capability was not merely the display of a bundle of skills; the deployment of the skills was intimately bound up with understanding of relevant underpinning knowledge; this meant that the claim to be able to distinguish levels of competence could not easily be dismissed. If assessors could distinguish levels of competence then they needed a range of grades to report different levels.

The draft code produced by the Working Group on Assessment covered a great deal more than grading scales (and aggregation formulae). However, the one main point that those interviewed latched on to was 'the uniform 20-point scale'. The range of responses is shown by the following quotes.

[A 20-point scale]..... is a nonsense and should be very strongly resisted. [1]

20-point grading is too fine, but competition is a part of life and we will have to settle for more than just two grades. [11]

I have substantial reservations about the degree of separation in a 1 to 20 scale. It gives a spurious impression of the accuracy with which assessment procedures can work [5]

A common assessment scale of 0 to 20 would not be difficult for written work, but it would be very difficult for clinical assessment. [3]

..... a 20 point scale from my perspective will be helpful. [9]

Focusing particularly on clinical assessment (and perhaps SSMs) there appeared to be a recognition that 'we sold the pass as soon as we accepted the case for a Merit grade'. If it was

thought that 3 grades were defensible and might be better for practical reasons than 2, then why not have four or seven or whatever? There were some signs of a growing consensus that in all clinical areas, and perhaps some others, people would eventually settle for a four-division grading scale. This would not make the purists happy, but it was practicable. Such a scale would be:

not-yet-competent; just competent; definitely competent; extremely competent

or unsatisfactory; just satisfactory; definitely satisfactory; extremely satisfactory.

This four-point scale could be 'sold' to students, but it was not the way in which final verdicts should be reported to the outside world. It was unlikely that any patient would be pleased to be treated by a physician or surgeon deemed to be only 'just competent'.

The writer cannot resist concluding this section with a lengthy verbatim quotation which illustrates some of the points made above. Those of a sensitive disposition should jump now to the next issue.

I know what is educationally correct. Students should be told they have reached the necessary level of competence and are safe to practise. Students want more than this because they are very competitive. And when people tell me that we mustn't encourage competitiveness, I retort that they are already competitive. In some places medical students are hacking into computer data-bases to get information on their level of performance; they are that desperate to know how they have done. This University is now coming down heavy that we have got to have grades. He wants a universal assessment scheme for the whole University; that awfully nice man – what's his name? I think that idea is plum crazy. Originally it was pass/fail and a very few distinctions. I told my students how they had done in the previous year and they said things like 'So I only passed by the skin of my teeth; if I'd known that I'd have worked harder'. So now we tell them if they've just passed and so the huge group in the middle is revolting, so maybe we'll split the 'definitely passed' into two lots. Come the revolution when we are forced to have twenty grades, why the hell don't we just give them their marks as percentages? Let's go back to what we did in my day. All this is farcical.....

[4]

(b) What do workload shifts mean for the management of assessment?

Assessment of medical students over the five years of the new curriculum involves a great deal of work for academic staff (and administrative staff as well). Much of the work necessary to ensure that assessment was effectively managed appeared to fall on the year co-ordinators for Years One, Two, Three, Four and Five, on the facilitators with a special responsibility for assessment and on the co-ordinator of SSMs. The people holding these posts who were interviewed did not **complain** about the burden of assessment they carried, but they were insistent on drawing it to the interviewer's attention.

It really is hard to get all the work done in the time available. There's a lot involved and we need more people on the ground to do the actual work. At the peak times of the year I have to pinch time from my other commitments. It can be a nightmare.

[10]

I'm not saying I have to devote too much time to assessment. But I do devote more time to it than I expected, partly because we have to think more about how we assess than we used to. But the practical and administrative burden is heavy too. We will

cope well enough this year, but next year when the full numbers in final year hit us will be purgatory on earth. [2]

240 students in each year; seven SSMs over the five years. Looking after the assessment is very hard work. You have to stimulate and co-ordinate the assessment thinking of all those who offer SSMs. Brief six external examiners and look after them. Get all the assessments in, recorded and published. There's a huge administrative burden – and a lack of administrative support. [11]

The new assessment arrangements required a great deal of work from people, but significantly these were not the same people asked to carry the responsibilities in the past. The new curriculum was not organised around the departmental delivery of specialised knowledge and as a consequence assessment was not now the responsibility of staff working under heads of department.

In the past various people in the department would do a number of lectures and then expect to contribute a number of questions to an exam paper, a number roughly proportional to the lectures they gave. Making up the exam paper was easy. Essentially the Head of Department just said, you, you and you, produce the questions.... [5]

It's different people who now carry the major burden of assessment. In general people within departments have less responsibility and involvement than they used to. The people who do most of the work are the staff of the Medical Education Unit and the year leaders. [8]

Different people, with different titles, in different places were seen as doing the work and inevitably this meant that different structures and patterns of responsibility were emerging; as one person said 'Academic power and authority lines have been broken and are re-forming'. People who had a co-ordinating role in assessment could not now require actions of people as could heads of department.

The workload of assessment has been centralised. I'm in the centre but it really is tricky to get other people on board. I need people's co-operation, but I can't order them to do anything. I have to ask 'please would you be willing to contribute a question?' and not just say 'contribute a question'. Authority is a difficulty; I have to rely on goodwill. We have built up a band of loyal helpers, but I am restricted in who I can approach. [6]

MEQs are integrated questions – integrated across blocks of PBL. I asked the block leaders to help; they said 'no, we'd be delighted if you did it yourself'. So I put the MEQs together myself, with considerable effort, not to say trauma, then hawk them around the block leader and subject experts. They don't find it necessary to contribute much – does that sound arrogant? – but they don't. They seem happy enough just to let me get on with it. [5]

I try to involve a lot of others, getting them to contribute bits of questions and mark bits of the exam. And I have to go down on my bended knees and beg them..... [4]

It was very clear that there was a lot of goodwill around; the general view was that 'people are usually helpful if you say please'. However, it was also clear that there were some people who were now unwilling to get involved with assessment; perhaps some of those who had a

reduced assessment input in the new curriculum welcomed their release and were getting used to devoting the extra time to their other commitments.

I've had NHS clinicians with honorary contracts with this university point blank refuse to help with assessment and this has probably meant that there are some parts of the course just not properly examined. And I'm not sure that anyone can do anything about it. The change in the curriculum has caused new problems in the management of assessment. [5]

There was an illuminating comment made that although some people welcomed being relieved of some of the 'hack-work' of assessment, they would not be so happy when they fully realised that they now had a diminished role in the high-profile aspects of final assessments.

We might still have problems when it becomes clear to some people.... let's say, just for illustration, senior figures in surgery, that they are no longer running their 'own' final exams. Touring round universities, not to say the good hotels, with senior colleagues doing the final vivas, has been of great significance to them..... [7]

Those interviewed did not suggest detailed changes in structures or in management that they wished to see. Perhaps possible ways forward can be inferred from the following three statements which are given in this case without attribution.

1. Assessment is now centralised; the Medical Education Unit is at the centre, but it is seen by some as a low-status body. Give it more status.
2. I suppose if someone were to foul-up in assessment, it is now the Associate Dean who would rap their knuckles.
3. We need a structure which directs assessment from the centre, with people there who lead and motivate and who are supported by a high-visibility review group.

(c) What kind of expertise do assessors need?

At the outset it should be clarified precisely why this is identified as an emerging issue. It is not that there is any judgement by this writer that those intimately involved with assessment within the new curriculum are in any way deficient in expertise, quite the reverse. The people interviewed were strongly committed to quality in assessment, they were tackling the challenges of assessment with impressive ability, ingenuity and thoughtfulness. This is an emerging issue because a majority of those interviewed had been prompted by the changes in assessment to consider just what it was that assessors should possess to do their job well. One person observed that in the past assessment was a thing to be done, not thought about: 'You want to test their knowledge – give them an essay. Are they any good clinically? I'm medically qualified, I can just tell...'

One contribution provides a particularly vivid perspective on the issue.

Because the new curriculum is centrally co-ordinated, people have become closely involved with assessment who have not gone through the traditional apprenticeship in the departments of medicine. Whereas previously young lecturers were inducted over a lengthy period into our assessment practices....well, now I do have some concerns. We have an excellent new Dean and Associate Dean who I think will be looking into

this... It has been very helpful for our people to talk with people like [the Convener of the Assessment Working Group]; he has become very well-informed about assessment in a way in which some of my colleagues frankly are not. He has spent a long time thinking about the principles and issues in assessment and our people, perhaps especially the honorary clinicians, have not. [7]

But what do these 'new' assessors have to say?

I know a fair bit about assessment and I've read quite widely. A lot of people in medicine don't have the foggiest idea; they don't see it as an academic area they should know about. So trying to persuade people to think about the ideas behind assessment is extremely difficult. The jargon alone may be enough to put some people off. I see committees working away on the curriculum and I don't know if there is enough assessment knowledge being put into them. There aren't many who know a lot about it, but there are some. And I'm not sure how much the knowledge there is about assessment is actually being shared and used. [6]

It may be remembered that in Section A(c) it was said that one person declined to be interviewed and for a particularly interesting reason. This person wrote: 'I have no specialised knowledge on the subject of student assessment. I doubt whether I would have anything to contribute to your research.' This prompts a number of questions which ought to be explored at greater length than is possible here. Is there 'specialised knowledge' of assessment that this person does not possess, but other people do? What is such knowledge like? What might it enable people to do that they cannot do without it? Should people who have a responsibility for assessment, but no specialised knowledge of it, be making an effort to acquire some? Why was it assumed that the researcher would wish to hear from those with specialised knowledge, rather than from people with important experience to reflect upon?

A few interviewees seemed to be saying that all that assessors need is experience. There is a hard-line response to this: experience provides an opportunity to learn but guarantees nothing. The gentler line would be that the experience gained through apprenticeship and length induction develops considerable 'craft-knowledge' within the assessment practitioner; this is probably adequate for proficient performance within an existing framework, but inadequate to deal with any radically new situation. If it is the hallmark of a professional that s/he is not only competent within a paradigm but critical of it, then this should apply equally to the practice of both medicine and assessment. If useful learning has to come out of the experience of assessing, then there has to be critical reflection on that experience. For critical reflection to occur, individuals need first to be convinced that it is necessary; then, most importantly, they need time to do it; they probably also need some knowledge input from others (or from the literature).

Despite all this about critical reflection on experience, a knowledge base for assessment does exist. Anyone laying claim to possession of that base should have acquired, through the scholarship of assessment, at least the following: knowledge of the language and central concepts of assessment; knowledge of relevant assessment policies; knowledge of the range of assessment methods available and of criteria for judging their applicability; knowledge of the research relating assessment to the quantity and quality of learning. However, the nature of that knowledge has to be appreciated; it does not as yet have sufficient theoretical 'respectability' and explanatory potency to prescribe what should be done in any particular situation.

Both medicine and assessment are complex professional activities gaining their character from an easily defined practical concern at their core; one could develop an extended analogy

between informed decision-making in medicine and assessment. Although the underpinning knowledge bases of medicine appear more robust and developed than those of assessment (in that more of them are scientific) it is probably no simpler to decide what treatment to prescribe for a patient than it is to prescribe how best to assess the learning of a student.

Medical teachers already need to know a great deal and they are under pressure from clinical commitments and the requirement that they be productive researchers. It seems unrealistic now to argue that they have to become both critically reflective of their experience in assessment and also assessment scholars, but some move in these directions seems inevitable if the claim to dual professionalism (as medical practitioners and as medical teachers) is to have any substance. At the least three things are needed: professional development, co-operation and time.

It was very striking from the interview data that although several members noted that they wished to learn more about assessment, only one person spoke of action that had been taken to help people learn.

Not everyone shares your fascination with assessment, Colin, but it is recognised here in our department – we're lucky. Dr A is on the GMC Assessment Group. Several of us are pretty familiar with the literature. But we did find that some of the tutors were not as good as they might be at providing constructive feedback to students. So we put on a training course for them..... [3]

It would appear that para 56 of the 1993 GMC Report needs to be given more attention. 'The changes in the assessment system will require considerable modification of the existing roles and practices of both internal and external examiners.guidance, if not training, will be required for those who examine in the new system.'

There is too much folk-wisdom and specialised knowledge of assessment for any one person to acquire. There must then be effective pooling of the wisdom and knowledge that does exist. This strongly indicates that there should be enhanced co-operation, partnership and communication amongst those who assess. There are, of course, obstacles to be overcome; the single most dispiriting remark in interviews with eleven positive and enthused people was 'In medicine there are still A-list people and B-list people; the As don't listen to the Bs'.

Individuals need adequate time not just to assess, but to develop their thinking about assessment. As assessment is so crucially important, it may be that time has to be **bought** to give more people more time to devote to the tasks of management, implementation and development of assessment. Where could such money come from? This goes beyond any knowledge this writer possesses. It may be fruitful to argue (a) that assessment is a vital part of teaching and (b) more of the funding available from the Additional Costs of Teaching purse be ring-fenced for assessment.

(d) Can assessment be both authentic and reliable?

This is another issue which deserves more extensive consideration than can be given here. The essential aspects of the issue have already been outlined. (See, for example, pages 18 and 24) Two quotations identify the gist of the matter.

There is an important issue here. Clinical assessment is authentic assessment because of its clear links with the realities of medicine, but there is a question about its reliability because of the variability across patients. So how do you cope with that variability without jettisoning authenticity? [2]

I see a tension between validity and reliability. We have put a tremendous emphasis on authenticity, in both clinically-situated MEQs and in clinical assessment. We have to move towards increased reliability, because we have to be able to demonstrate the reliability of our assessments in the face of increasingly demanding students. But we can't move too far towards reliability, because you can't do reliable assessment on real patients. The USA National Licensing body depends purely on paper-based tests, because of the pressure for reliability.... I am uneasy..... [3]

The traditional relationship between validity and reliability is simply expressed: reliability is a necessary but insufficient for validity. In other words, a reliable assessment can be invalid, but a valid assessment cannot be unreliable. This has sometimes had the effect in the past of people going all-out for reliability first and then adding on a concern for validity almost as an afterthought. More recent writing on assessment has tended to de-couple the two concepts and has recommended that the designers of assessment aim first for maximum validity (often re-christened authenticity) and then try to make those assessments as reliable as they can be given practical realities and resource constraints. There has also been a widespread movement recently for educators to emphasis validity and for administrators to insist on reliability (because they often face the brunt of student appeals). Those who see assessment primarily in terms of student learning prefer validity; those who see it primarily in terms of accountability prefer reliability.

How then is maximum validity to be achieved? Validity (the extent to which an assessment actually does what it claims to do) is increased by a continuing concern for the alignment of course purposes (and what students do after completing the course) with the actual nature of the assessment instruments. Is this assessment valid in terms of what we say students should be learning? This initial concern for face-validity can of course be supplemented by rather more technical measures and research evidence.

How is optimum reliability to be achieved? The single most important activity is the clarification and explication of the assessment criteria. This is not enough, as one can have explicit criteria to which different assessors attach very different meanings. The meaning of assessment criteria can be made more obvious via model answers, detailed marking schemes and so forth. This too is not enough. There is sound research evidence which shows that assessors only reach acceptable levels of consistency of judgement when they are members of an effectively inter-communicating group. (This is very obvious when staff talk of 'inducting new assessors into our thinking' and 'training a new external examiner in our ways'.)

In the responses of interviewees there was substantial evidence provided of a committed search for authenticity and validity and of a determined effort to make assessment criteria more explicit to all staff (and to students) and to encourage shared understanding of their meaning. There were very striking examples of assessors communicating with each other (for example, in relation to the MILE); nevertheless several people said quite definitely that there was insufficient communication amongst assessors, both in the construction of assessment instruments and in the marking/grading of student products and performance. Communication occurred within pockets, but not between pockets. There was a wish for communication and co-operation, but the usual enemy was identified – a lack of time.

On page 24 a simple conclusion was stated; it seems appropriate to repeat it. There is reliable assessment which is quite cheap; there is authentic assessment which is quite cheap; assessment which is both reliable and authentic does not come cheap. Money may have to be found to purchase time for assessors to do the high quality job they want to do and which their students deserve.

H. IMPORTED, GENERIC ISSUES

In this penultimate section, five generic issues are addressed. They differ from those identified in the previous section in that the researcher had them in mind before any interviews were carried out. They are issues which it is predicted would be of relevance in all the four subject areas to be studied. There were again no direct questions on these matters in the original schedules. When people raised these matters, questions of clarification and amplification were asked. In what follows the issue is first defined and then a brief summary is given of main points from the data. This report is already too long so illustrative quotations are minimal. An apology is offered for the attack of compulsive alliteratitus ('The CAs') in issues two to five.

(a) Assessment: The Big Purposes

In the literature, assessment is sometimes described as having three broad purposes: certification, accountability and learning enhancement. At other times, the first two are conflated and two broad purposes are given: certification/accountability (the control function) and learning enhancement (the growth function).

There was not much reference to the accountability purpose i.e. that assessment produced evidence that courses were doing what was claimed for them and that individual members of staff were doing their job as teachers effectively. One person noted that more use could be made of assessment data in course evaluation; another noted that assessment purposes carried out by the PBL facilitators with their groups provided them with feedback on their performance.

Everyone clearly acknowledged the certification and learning enhancement purposes; both were seen as inescapable. Although there was a little concern about the 'rigour' of clinical assessment in the final years, there was no fear that the qualification awarded would be less deserved or have lower currency. In other words, the assessment arrangements in place for the new curriculum would preserve standards within the Medical Faculty. There was general agreement that the new assessment scheme did more to enhance student learning than had happened in the past: 'Assessments are now in place that encourage students to learn better'. However, the number of references to giving more attention to formative assessment suggests that people are aware of plenty of scope to fulfil the growth function even better.

(b) Constructive alignment

Constructive alignment is a concept which has appeared only recently in texts on assessment. It refers to that situation in which the aims and objectives of a course (perhaps described as statements of learning outcomes), the methods adopted for teaching the course and the ways in which learning is assessed are all planned together and related to each other so that each guides and reinforces the others. Outcomes, methods and assessment then co-exist in a positive, productive relationship i.e. they are constructively aligned.

Assessment procedures had been planned as part of the overall process of designing a radically new curriculum; course approval procedures had required the designers to demonstrate that purposes and assessment processes were adequately aligned. It is thus not very surprising that the majority of people interviewed made some reference to the ideas involved in constructive alignment and to the fact that it had operated as a kind of design principle. What surprised the interviewer a little was that the actual term was used by three interviewees; it was never used in the other three subject areas explored. What might be seen

as a recent coinage, a new piece of specialised terminology or an unwelcome bit of jargon had become for some an unremarkable part of everyday language usage.

(c) Communicating about assessment

This issue has already been introduced into the discussion of how optimum reliability in assessment can be achieved. Clear assessment criteria are needed for consistency in assessment but alone they are not enough. What is needed in addition is shared understanding of what the criteria actually mean. Although things like mark schemes and model answers help to attach meaning to criteria, shared understanding commonly comes about by assessors discussing, and indeed arguing, amongst themselves about how criteria are to be applied to actual examples of student work and performance.

There is a familiar, and no doubt idealised, picture of how assessment used to take place within conventional academic subject departments. Staff knew their students well; exam questions and papers were 'shredded' in group meetings; new members of staff and external examiners were inducted into 'how we think around here'; the second marking of essays and theses led to animated debates in coffee rooms and corridors; meetings of examiners were lengthy affairs with time for extended discussion of individual cases. Such a system was not always in a student's best interests and no doubt was open to abuse, but one obvious thing about it was the number of opportunities it provided for assessors to communicate with each other. It is now common-place for assessment to occur by a process which has none of the above features.

In the interviews with medical staff much was said about the efforts that had been made to provide clear and explicit assessment criteria; rather less was said about co-operation, collaboration and communication amongst assessors. For example with MEQs it was normal for them to be produced by one person, who would then send them out for e-mail comment by people who often did not reply; marking of these was then normally done by one examiner marking all of one part for all students. This latter meant that although the consistency of judgement of one examiner over time might be an issue, inter-assessor reliability was not i.e. the need for inter-assessor communication was reduced. One person noted that it was 'difficult to generate any team feeling amongst assessors on a big multi-disciplinary integrated exam'.

However three examples were spontaneously provided of occasions where assessors were said to communicate very effectively about assessment (there may well be others not mentioned). Firstly there was the MILE (see page 5); secondly there was the LCP (in which there had to be much communication amongst University staff and general practitioners); thirdly there were termly discussions amongst facilitators and block-leaders about student progress.

Constraints on communication were noted: 'we are talking here about a lot of very busy people who obviously can't waste time on wee social chats' and 'this is a very dispersed exercise –communication is OK for us here at the centre in the Medical Education Unit – but there's a lot of others out there'.

Assessment is a crucial activity that requires to be conducted with humanity and with professionalism. It may have to be promoted up the list of priorities if it is to be seen as an activity for which people must find time. Talking about assessment should not be dismissed as a trivial activity; it may be that there is a central role here for the Faculty or the Medical Education Unit in ensuring purposeful communication about assessment. This relates to what was said earlier by those interviewed about the need for an assessment review group.

(d) Criterion-referenced assessment

The distinction between norm-referenced assessment ('N-R A' - spreading candidates out to make judgements about their relative standing) and criterion-referenced assessment ('C-R A' - judging candidates' absolute standing against predetermined criteria) is a familiar one of long-standing. It is important to ask what these things are. They probably represent polar extreme schools of thinking about assessment or even ideologies to which people claim allegiance in an uncritical way. It is doubtful if they can ever be used accurately to describe any actual assessment scheme or strategy. In practice, pure forms of norm-referenced and criterion-referenced assessment do not exist.

Everything written about appropriate grading scales in Section G (a) could now be re-interpreted as a conflict between criterion-referenced and norm-referenced thinking. The situation described therein can be seen as people trying to think their way through to a workable compromise between conflicting ideologies of assessment. There appear to be two different dimensions in medicine along which people vary. Firstly administrators, or people thinking administratively, prefer students to be spread out (N-R A); 'rank-ordering of students helps with appeals'; educators, or people thinking educationally, want most students to be clearly successful (and a few clearly unsuccessful) (C-R A). Secondly, some see clear differences in the levels of academic understanding that students achieve and thus a justification for differentiating students across a spectrum of grades (N-R A); others think in terms of a binary distinction between competent and not-yet-competent. (C-R A).

At the time of writing it looks as if the Medical Faculty may agree to having a University scale to spread students out in 'academic' matters and a binary decision to report on clinical assessment. It is interesting to note that the GMC option of using SSMs to identify outstanding achievement was in fact used to assist with the award of honours and distinctions, although it did not feature in any contributions from the interviewees.

(e) Constrained autonomy

The idea of any single academic having complete freedom to do what s/he thought best in assessing students is, and always has been, absurd; individual autonomy is always constrained. The idea of a group of academic staff in one subject area having freedom with respect to how 'their' students are assessed is not self-evidently ridiculous. However, even here clear constraints operate and are recognised. Students will be awarded a degree not by the medical faculty, but by The University; thus The University clearly has to exercise its legitimate authority over what is done in its name. This manifestly includes how results are recorded, aggregated and conveyed to a central registry, but it also includes aspects of quality control.

The existence of such constraints on individuals and the faculty were seen as entirely proper and reasonable, even when details of how they emerged and were exercised were sometimes criticised. However, people did have worries about how their autonomy as medical teachers might be increasingly constrained in the future. Examples were given of how constraints might be extended into new areas of activity (e.g. grading scales), might increasingly come from supra-institutional sources (e.g. the Government, the QAA), and might be imposed in a more directive way (e.g. the GMC might prescribe rather than advise).

As with the other subject areas external influence was not seen as malign when it came from people perceived as having the appropriate subject-based experience and credentials. Pressure from the GMC would be 'taken very seriously into consideration – after all they are

themselves medics'; pressure from The University was regarded with more suspicion because 'they tend not to appreciate our distinctive character'.

I. CONCLUDING REMARKS

In the opening paragraph three research questions were identified. Why do you assess students in the ways you do? When assessment practice changes, why does it change? What assessment issues emerge as important? Brief answers to these questions, summarising the views of the interviewees, will now be attempted.

- Medical students are assessed within the new curriculum in the ways described in Section B. This assessment system was seen as significantly different from the previous one.
- Assessment procedures were changed because the curriculum as a whole had been re-designed. The 1993 GMC Report (*Tomorrow's Doctors*) was seen as very influential. There was also a general feeling that change in assessment had been needed and concern for the quality of student learning had pointed to the forms it might take.
- Five features of the new assessment system were regularly said to be good. The assessment system is workable and it has worked well so far; student and staff reactions have been positive; the MILE, the OSCEs and the MEQs are particularly effective methods of assessment; there is a pleasing diversity and balance in the assessment methods deployed; there is more integrative assessment; the assessment scheme as a whole is coherent.
- There were said to have been four major changes in assessment: less assessment of factual knowledge; more integrated assessment; more standardised clinical assessment and a reduced dependence on 'traditional' assessment methods.
- There were recommendations for further change. Most of these related to three broad themes: giving students a better idea of where they were; providing a better review of assessment as a whole; further improving clinical assessment. There was some concern about assessment in Year Four particularly.
- Four issues emerged as salient: the appropriateness of different grading systems; the management of workload shifts in assessment; the kinds of expertise that assessors need; and the tension between authenticity and reliability in assessment.
- The abiding impression left with the researcher was of a group of people of goodwill and with a real commitment to operating an assessment system of quality. They did not, however, seem to have adequate time to think about assessment, to communicate amongst themselves and with others and to initiate further developments. Interviewees saw implications here both for the future review and management of assessment and for the ongoing professional development of those who assess.

ASSESSMENT IN HIGHER EDUCATION

Subject Area IV

Design: a descriptive report

Version 2

A. SETTING THE SCENE**(a) The Research Purposes**

The research aims to illuminate the underlying rationale for current practice in assessing student learning; it aims for description and explanation only. The researcher has no role as an evaluator, although the evaluations of others are described. This phase of the research is guided by three general research questions: Why do we assess students in the ways we do? When assessment practice changes, why does it change? What issues in assessment emerge as important?

(b) The Course(s)

The focus of the research was the assessment of student learning in the general subject area of 'Design'. Students of design spend much of their time in the School of Design and Craft; this is one of three schools within The Glasgow School of Art, which is itself an associated college of The University of Glasgow. It was decided that the research would concentrate on the assessment of students following courses leading to the degrees of Bachelor of Arts (Design), Bachelor of Design (BDes) and Master of European Design MEDES (Product Design) and Bachelor of Engineering (BEng) and Master of Engineering (MEng) (Product Design Engineering).

The School of Design and Craft (SODAC) has seven component departments. Five of these (Ceramics, Interior Design, Silversmithing and Jewellery, Textiles and Visual Communication) deal with students seeking the degree of BA (Hons) Design. The Product Design stream for this degree ends in June 2002 and is being replaced by two first degree courses, the BDes (4 years) and the MEDES (5 years); the first cohort of students seeking either of these two degrees is, at the time of writing, at the end of the second year of study. Students of Product Design Engineering divide their time between The University of Glasgow and SODAC. The more theoretical and academic aspects of the course are the prime concern of the Engineering Faculty (Glasgow University); all of the more applied aspects of studio design practice are the responsibility of the Department of Product Design Engineering (Glasgow School of Art). Degrees awarded are the BEng (4years) and the MEng (5 years).

All students on the BA (Hons) Design programme follow a common first year course (for which there is a First Year Co-ordinator). In years 2, 3 and 4 they are based within one of the five specialist departments. In addition some time in all years is devoted to Historical and Critical Studies; in this component students work with those from other design specialisms and from Fine Art. Given the range of provision, the changing nature of provision and the year-on-year variation in student numbers, it is not easy to give a simple summary indication of the scale of the design operation. Approximate student numbers are as follows.

Average number of students per year	Ceramics	14
	Interior Design	10
	Silversmithing/Jewellery	14
	Textiles	36
	Visual Communication	40
	Product Design	22
	Product Design Engineering	35
	Historical and Critical Studies	171

(c) The People Interviewed

Approval for the research was obtained from the Chair, School of Design and Craft. She was then asked to suggest the names of about eight members of staff whom it was recommended should be approached for interview. After discussion, nine names were provided.

These nine people were approached by letter in February 2001; the letter covered the purposes of the research, the main topic areas for interview and an indication of the time that might be required. Enclosed was a protocol making explicit the 'rules of engagement' and covering aspects of approval and anonymity. The letter was followed up by an e-mail message one week later. Within a very short period after the e-mail, nine positive responses were received and dates and times arranged. This was the only one of the four subject areas in which no-one declined to take part.

The nine people interviewed were the heads of the seven component departments of SODAC, plus the first year co-ordinator and the head of Historical and Critical Studies. All of those involved were experienced as assessors of student work; the length of experience within Glasgow School of Art varied from five years to over thirty; everyone had some experience of assessing students in other broadly similar institutions, whether in the UK or elsewhere in Europe or in the USA. It has to be noted that, due to the particular structure of SODAC, these nine were not directly comparable with those in the other subject areas. All nine were responsible for some aspects of assessment system design and management in addition to being heavily involved with the day-to-day business of student assessment. In the other groups of interviewees, there were people who were assessors, but who had no such role as 'assessment managers'. It is **not** claimed that the nine form any kind of representative sample of the total population of those who teach and assess design students in Glasgow School of Art. Rather they form a purposive sample, chosen because they satisfy the criterion of being well-qualified, through a variety of types of experience and involvement, to speak on the topics in the interview. There is no intention to make any empirical generalisation beyond this group.

As a consequence of the structure of SODAC and the composition of the group of interviewees, it has proved more difficult in this subject area to make the sources of views non-identifiable. It is of course easy not to name people.

With each person's permission, all interviews were recorded on audio-tape. Interviews lasted on average for 59 minutes (excluding preamble and farewells) and ranged in length from 38 to 75 minutes. It is interesting that in several cases the interviewer was taken on a tour of the studio spaces and introduced to students at work; he was also regularly offered coffee (and in one instance a Fry's Chocolate Cream). Such experiences were not offered in other subject areas. The researcher appreciates the generosity of design staff and thanks them for it.

(d) Design as a Subject Area

In choosing design as one of the subject areas to be explored, there was an assumption that it made sense to treat design as one subject area. This assumption was confirmed in interviews as being justifiable, as long as it was recognised that within one very broad common purpose and a lot of shared assumptions and language, there were clear differences amongst 'disciplines' and indeed within them. (There were different emphases, for example, across Graphic Design, Illustration and Photography although together they made up Visual Communication.). The fact that there was a school of design at all reflected the general belief that there was a general concept of 'design'.

It does make sense to think, broadly, of design. We share a lot of words, categories and ideas across the departments – even although different departments are all doing quite different things.

There's quite a lot of similarities in what we do. It's OK to think of design as one subject area. And there's a shared ethos within the school....

(e) 'The Four Quadrants' Where is Design?

Interviewees were given a copy of the diagram above; it was introduced like this. 'It is said that subject areas are differently located on this diagram. The axes represent what is emphasised in assessment near the end of a degree course. Please do two things for me, (a) comment on the framework itself and (b) place design where you think most appropriate.'

The first point to be made is that, whatever reservations people had about this simple diagram, almost everyone had a lot to say about it. Invited to comment on the relevance of this theoretical framework to their subject area, these teachers responded with relish. The overall reaction to the content of the framework was rather less positive than in the other three areas; there was, however, the same tendency as elsewhere to explicate what the design curriculum **should** be like rather than to describe what was actually emphasised in final year assessments. It seems likely that in this design there is a genuinely close correspondence between what the curriculum is said to be like and what actually gets assessed i.e. between the rhetoric of course purposes and the reality of assessment practice. The correspondence is probable given the centrality of portfolio assessment, however confirmatory evidence would require not only analysis of what is sought in assessment procedures but also of what it is that actually gets rewarded in student performance. This goes well beyond the present research.

The most frequent comments made were as follows: (a) fitness to practise does get emphasised, but not in any narrowly vocational sense - students are not assessed for their readiness to enter any one particular career; (b) the different disciplines within design would probably be located differently, historical/critical studies would be different from studio practice, engineering would be different from studio practice in product design engineering; (c) understanding of concepts and evaluation of alternatives are not polar opposites, they should proceed in parallel. Interestingly two people offered alternative frameworks which

they thought better captured the situation existing within design; although their use would provide effective illustrations of the differing characteristics of specialisms within design, they did not seem to this researcher to focus on key differences in what gets assessed in different subject areas.

When it came to locating design on the diagram, people fell into two main groups. One group placed design firmly in the upper right quadrant (the one that has been described as 'applied and soft'); one group said that design had to embody both sides of both dimensions, thus it could only sensibly be depicted as a shape centred on the origin (0,0). One person's response was completely different to all the others: s/he wished design to be placed right at the bottom of the vertical axis (0, -4) because design had to be all about developing creative individual people to their full, highly varied, potential. The course purposes were educational rather than vocational – pure rather than applied. (Note, however, that this is one of those statements about desired course purposes rather than existing assessment realities.) If these important differences are temporarily set aside and a crude average for all the design interviewees is calculated, then design is located in the upper right quadrant, (at +1.5, + 2.0). This represents a perception that in the final assessment of design students, rather more emphasis is placed on their fitness to practise than on their academic soundness and rather more on their ability to evaluate alternatives than on their understanding of a bundle of defined core concepts. Design was chosen as one subject area for this research in the hope that design teachers would locate it in the upper right quadrant; this, on average, was what they did.

B. STUDENT ASSESSMENT: THE PRESENT

(a) How are students currently assessed?

This report is being written with two distinct audiences in mind: the staff of the School of Design and Craft and a wider group of people interested in assessment in higher education. The former are very knowledgeable about their own subjects, courses and methods of assessment; the latter are likely to know very little about design, how it is taught and how it is assessed. There is a real danger that this section will be very long indeed and contain much that is unnecessary for one or other of the two audiences. In the interests of economy an attempt will be made to organise masses of material under three headings: (i) clarification of key features common to most design courses; (ii) the general pattern of assessment of studio practice within BA (Hons) Design courses in five specialist areas (Ceramics, Interior Design, Silversmithing/Jewellery, Textiles and Visual communication); (iii) the main differences in assessment in two remaining areas (Product Design and Product Design Engineering) and (iv) assessment of Historical and Critical Studies. This treatment will inevitably gloss over many points of detail which may be seen as significant by individuals and the necessary selection of material may well produce distortions of emphasis.

(i) Clarification of key features

Portfolio assessment

The idea of a portfolio of work has been familiar in such fields as art, architecture and design for a very long time; the idea of portfolio assessment within education is relatively recent, but it has extended well beyond its original areas of application to the assessment of practice in a number of fields and at a variety of levels. To give just one example, the accreditation of higher education staff is achieved through the submission of a portfolio of evidence to support the claim of capability as a successful, professional teacher.

Originally a folio was a sheet of paper and a portfolio was a case to contain several sheets. Although the familiar flat, black portfolio is still a common sight in art schools, the meaning of the term now extends well beyond this. A portfolio could be a box, a shelf or a cupboard containing student products of one kind or another; it might have no clear physical container at all but be better thought of as a collection; it might have no obvious physical reality and exist only as a virtual 'e-collection' in a computer hard-disk, or network.

All portfolio assessment, and not just that of design students, has itself to be 'designed'. The portfolio for assessment might contain only 'final' products (say ceramic pots or textile wall-hangings) or it might contain supplementary materials to provide evidence about the processes by which the final product was achieved or about understanding of the context in which they should be located. As the portfolio is unlikely to contain **all** the work produced by a student, what are to be the principles of selection? What guide to the portfolio contents does the student supply? (Is such a guide an index, a map, or a set of claims with directions to the appropriate evidence?) Should earlier assessment reports be included? Or the reports/testimonials of workplace supervisors or employers? Is it required that the portfolio be easily adapted for different target audiences? Should there be any difference between a 'developmental portfolio' and a 'showcase portfolio'?

The contents and organisation of the portfolio are likely to be strongly influenced by the assessment criteria which are to be applied to it. These criteria should be explicit, transparent and, as far as possible, the subject of understanding shared by the student and any significant assessor.

The people interviewed seemed to be aware of all the above issues. Just one quotation at this point...

All of us, I think, use portfolio assessment in a broad sense. The portfolio contains both the products created and the back-up material to show how they were created. We make it clear what the 'deliverables' should be: log-books, concept sketches, informal and formal drawings, presentation drawings, models (both developmental and final), test rigs to test structures and so on. Understanding of the criteria by which the portfolio will be assessed is gradually built up over the years of the course.

One should also ask if the assessment of studio practice is based only on the portfolio itself i.e. does the portfolio contain all the evidence to be taken into account by the assessor? What is clearly not present within the portfolio is evidence which comes from the student **talking** about his/her work, either in an oral presentation or in response to questions from an assessor in both formal and informal settings.

Criteria and levels

Much of the assessment of studio practice within all the design courses made use of three central criteria. Each broad criterion was subdivided into five or six sub-categories; these are indicated below. (There were some variations in the sub-categories across different specialisms.)

¹ The principles and procedures of portfolio assessment are dealt with by Heywood (2000, Chapter 12). *Assessment in Higher Education*, London: Jessica Kingsley Publishers.

Study and Investigation

- Analysis of project briefs
- Assessment of information required
- Information gathering (primary and secondary sources)
- Prioritising of information
- Materials investigation
- Personal initiative demonstrated

Creative and Imaginative Development

- Divergent thinking
- Development of alternative ideas
- Critical analysis
[Selectivity]
- Refinement of ideas
- Studio/workshop testing of ideas

Professional and Executive Ability

- Project execution/organisation
[Appropriateness of ideas]
- Presentation skills: verbal and visual
- Craft skills (technical competence)
[Group participation]
- Time management

In year one, there were four main criteria rather than three; the fourth was 'studentship' and covered things like attendance and time-keeping. In subsequent years students were assumed to have internalised these; however, the requirements might still appear either as a sub-category of Professional and Executive Ability or as an additional separate requirement, a hurdle which students must clear.

Performance on each criterion was normally graded at five levels: excellent, proficient, competent, basic or unsatisfactory. There were complex procedures for the later translation of these levels into letter grades (A to E), percentage marks and degree classes (1, 2.1, 2.2 and 3).

An 'assessment grid' (of levels against criteria) was used by tutors and assessors in both formal and informal assessment of students; its use was also encouraged and valued for developing student abilities in self-assessment.

If they are detached from the contexts in which they were used, these assessment criteria invite many questions and critical analysis. Are the three broad criteria of equal importance? Within each broad criterion, what weightings are given to the component sub-criteria? What measure of compensation is allowed within and between categories? Does the ordering of sub-criteria suggest a chronological sequence of activities or a hierarchical ordering of skills – or both? Are all alternative ideas to be developed before any critical analysis of them? Why does group-working feature in some specialisms and not others? What conception of professionalism underpins Professional Ability? And many more.....

It was regularly said within the interviews that students encountered these criteria early in the course and lived with them throughout their course, gradually coming to understand each criterion and the relationships between them. Staff went to considerable lengths to produce helpful, illustrative documentation and to keep referring to, and discussing, the criteria in tutorial sessions and at the workbench. It was frequently claimed that all students and staff

came to a shared understanding of what the criteria meant, what meanings should be attached to them. It was admitted, however, that shared understandings could be local to one specialist area and that different meanings developed in different areas. The framework was, more or less, the same throughout the School of Design; this did not mean that it was similarly interpreted in all areas.

We publish the criteria; they are open and transparent right from the beginning. We keep showing and telling students what they mean and a good understanding of them develops throughout the course. There are feedback tutorials to pick up concerns and to clear-up misunderstandings.

Assessment is against the same criteria all the way through; students are getting used to them from the time they come in the door. All the Art School works to roughly the same criteria. And note how our criteria describe the important features of the design process. And we use the same criteria when we ask students to do a self-assessment....

As long as I feel the words are relevant to us, that's fine. I know we can interpret them as we want to within our particular specialism. So I have no problem with the way the criteria are written down.

There was one matter which was not made explicit in the interviews; perhaps the interviewer failed to realise it should be. Were judgements of excellent, proficient, competent and basic made in relation to the standards expected at that stage or in relation to the standards it was hoped that students would reach by the end of the degree programme? Are all four levels equally open to students in all four years? There may well be a temptation, even when there is no intention, to think that in year 1 students should normally be judged as basic, in year 2 as competent, in year 3 as proficient and in year 4 as excellent. This would obviously have the cynical advantage of demonstrating progression through the programme. Probably it was assumed to be obvious that levels of performance were to be judged in relation to what could reasonably be expected of students at a given stage; were this not so there would be little distribution of students across the final degree classifications.

Interim and final assessments

Informal assessment of students clearly took place throughout the year. Staff communicated frequently with students as they worked in the studio and in regular tutorials; inevitably this involved that informal assessment which is integral to, and inseparable from, interactive teaching and learning. In most cases, in all years, it was usual for there to be two more formal assessments. These were the Interim Assessment (which took place towards the end of term 2 in, say, March) and the Final Assessment (which took place at the end of term 3 in June).

The purpose of the Interim Assessment was diagnostic and formative. It was to give students an indication of the standard achieved in each aspect of their work thus far, a description of their strengths and weaknesses and appropriate guidance as to what should be done to enhance their final Assessment. Interim and Final Assessments were similar in nature and both were informed and structured by the 'Assessment Grid'.

There was some variation in how staff perceived student involvement in the Interim Assessment. In some areas staff saw themselves as the sole assessors, reporting their assessments to students and discussing their verdicts in the light of the assessment criteria. In other areas, students were asked to assess themselves, then staff assessed the students and then openly explained to students any differences between the results of self-assessment and

tutor-assessment. In this latter case student self-assessment might influence the assessment by the tutor; in the former case it could not. In neither case was responsibility for any component delegated to students.

In the interim assessments students do a self-assessment exercise on the same set of criteria. Staff provide the actual assessments and then discuss any discrepancies between their judgements and those of the students. I find this is a very useful exercise and it can be very revealing to both students and ourselves. I find that the brighter students tend to be very hard on themselves and conversely the less creative often have an over-inflated idea of their abilities.

Showing, submitting and showing

As students work in the studio, staff are routinely involved in moving around observing, commenting and discussing; learning, teaching and assessment are all taking place. Only observational research could reveal what it is that tutors actually say in these encounters; some interview data suggest that, apart from questions, tutors characteristically issue 'gentle imperatives' such as 'Tell me what you are doing...', 'Explain why you are doing X rather than Y...' and 'Show me...' This last mentioned applies to both processes (seeking the demonstration of the use of skills) and products (seeking display of artefacts, completed or in construction). It is worth stressing that this availability of direct observable evidence is a central feature of design. The position is very different in the other subject areas in this research. Clinical skills/processes are increasingly important in medicine, but observable products are relatively rare (perhaps a sutured wound, a completed prescription form?); practical skills and products in chemistry are only important in laboratory work; there is no obvious parallel in philosophy, perhaps the closest approach being the production of oral evidence in the generation of spoken arguments.

When it comes to the evidence for formal assessment, the design student supplies this via the portfolio. The student has some choice in what s/he submits for assessment by those recognised as having a legitimate role as assessors. In, for example, the final year, the student prepares the 'degree submission'. This includes everything to be taken into account by the internal and external examiners. This collection is then shown/exhibited to the assessors.

In design (as in fine art) there is a further phase beyond the degree submission and that is the 'degree show', a public event that takes place very soon after the formal procedures of final assessment. These degree shows are of considerable interest to a wider public and usually attract the attention of critics from the media.

The degree show is basically the same stuff as in the degree submission. They put up a display of what they think is their most important work; some of the preparatory and contextual stuff is taken away before the public sees it – but most of the degree submission is seen by the public.

It may be questioned whether the degree show is relevant in a discussion of student assessment. It is generally accepted that such shows have a profound, positive motivational effect on students: 'Knowing there will be a public showing is a strong motivator on everyone; some students produce work of quite exceptional and unexpected quality'. The show affects standards; there can be little doubt that it also has an influence on the meanings attached by staff to the criteria applied in the assessment of design products. Staff thinking on assessment criteria is commonly challenged and informed by the views of the wider audience. From a student point of view, the degree show cannot be described as formative in any conventional sense, coming as it does after the final assessment is completed; it may well,

however, be formative in that public assessment may confirm or challenge staff thinking on assessment and may influence students' future development as career designers.

Describing this aspect of assessment as 'show-submit-show' omits another important feature of some courses. Exhibitions of student work are occasionally arranged in mid-course. These could be special one-off displays somewhere within the art school or they might be externally mounted for a special target audience, for example of industrialists. Clearly these can have a powerful formative influence on students; they also inform staff perceptions of what is seen by other people as excellent, good and unsatisfactory in the design field.

(ii) The general pattern of assessment within studio programmes for BA (Hons) Design

First Year Design is a full-time course of study; each week of the session involves four days of studio work, the fifth being spent on Historical and Critical Studies. All the design and craft specialist areas are covered, helping students to develop a wide range of practical and intellectual skills and to confirm or revise their initial choice of specialist area.

In the studio course, students complete a series of projects, supported by practical workshop sessions, seminars, critiques and lectures by working artists and designers. As the year progresses the projects become increasingly student-defined. The four assessment criteria described earlier apply; these reduce to three as 'studentship' becomes subsumed as an aspect of Professional Ability. There is a progress review (end of term 1), an interim assessment (end of term 2) and a final assessment. At this last, work from the student's portfolio is exhibited 'on-the-wall' and the actual assessment is carried out by all first year design staff, supplemented by staff from SODAC, from Fine Art, from Historical and Critical Studies and by an External Adviser. The different assessors each complete an assessment form; the student is thereafter presented with one consensus form, conflating the different assessments. Every student then has a feedback tutorial in which the tutor deals with any concerns or misunderstandings.

Each student is graded Pass or Fail on each of the assessment criteria and requires a single overall Pass grade to proceed to the next year. It should be noted that the final assessment decisions are simple binary ones between 'yes' and 'no' and that no numbers/marks/grades are used.

Assessment of studio practice in years 2,3 and 4 has many of the same features. For example, students work on a series of projects which build into a portfolio; the same three broad assessment criteria are used all the way through; in each year there is an interim and a final assessment; an increasing range of assessors is involved in addition to the tutor; assessments are not cumulative i.e. the assessment at the end of each year matters only for progress, the Year 4 assessment, is the only one which determines the class of degree awarded; external examiners (one per specialism) meet and talk with students as they work and not just for a final oral; student work is exhibited at various times, most importantly in the degree show which comes very soon after assessment of the degree submission.

The number, length and nature of projects varies across the specialisms and from year to year. In very general terms, the projects become longer, more student defined and directed, and more contextualised as the degree programme unfolds.

Two features were stressed by all the five interviewees responsible for BA studio-work assessment: firstly, assessment takes place within a context of session-long interaction with students both as they work 'at the bench' and in regular tutorials and, secondly, assessment is

authentic – students of design are assessed in ways similar to how design practitioners are evaluated in the ‘real’ world.

Assessment is not a detached thing; it's going on in the tutorial meetings and we have one of these per week. We are closely in touch with students as they progress through a project, we are discussing and monitoring their work all the time. And then when it comes to putting an exhibition together, helping them find their distinctive voice. And remember the exhibition is real; it's not just a plan or an intention. It's a major event in their lives and it's the bit that really counts. At the end of the course you're only as good as your final assessment.

All areas used similar assessment criteria; use of these was encouraged in all informal assessment by tutors and in self-assessment. All areas were required in their final assessment decisions to award each student with a particular class of degree. However, the procedures adopted for moving from descriptive assessment to degree classification seemed to vary considerably. It appeared to depend on the general attitude of heads of department to numbers. Percentage marks, it was said, either gave a dangerous and spurious appearance of precision, wholly inappropriate to design, or they allowed useful and sensitive distinctions to be made between students similar in achievement.

(iii) Assessment of Historical and Critical Studies

All students in the School of Design (and in the School of Fine Art) are in continuous contact with the Department of Historical and Critical Studies throughout their course. Students study the history and theory of art and design, outwith specialism-based boundaries, ‘in order that they are better able to see their specialisms in a broad context and to inform their own work accordingly’. Satisfactory attendance is required of students at a varied programme (which has both mandatory and optional elements) of lectures, discussion groups and seminars. There are guides to reading and reference materials and progress throughout each session is supported by regular individual tutorials with personal tutors.

The main methods of assessment are essay, oral presentations and reports of work-based learning. The exact programme of assessment depends on the choice of options pursued. Typically in Year 1 a student would submit two essays and deliver one presentation; in Year 2 there might be three essays and one placement report. (Placement might, for example, be in schools, hospitals, museums or galleries.) For every assessment students are provided with a check-list (‘Have you attended to these things?’); such check-lists act as both aids to preparation and as a prompt to self-assessment; the elements in the check-lists are explicitly related to the assessment criteria which will be applied by the assessors. Each assessment is given a summary grading, A to E; the component assessments are then aggregated to a single grade at the end of a session; a passing grade is necessary to progress to the next year of the course.

In Year 3 the main piece of work assessed is the Part 1 Essay. In consultation with tutors, Students elect to develop this either into an Extended Essay (to be presented on enrolment for the final year) or into a dissertation (to be presented at the beginning of term 2 of the final year). An extended essay accounts for 15% of the total marks determining the final class of degree award; the dissertation counts for double i.e. 30%. Extended essays and dissertations are read by two or three internal examiners; most of them then go on to an external examiner as well.

(iv) The main features of assessment in product design and product design engineering

Product Design

The product design stream within the BA(Hons)Design programme will end in June 2002. Students in years 3 and 4 are currently being assessed in the ways described above. Students of product design in years 1 and 2 are following a programme leading to either BDes (in 4 years) or MEDes (Master of European Design – 5 years). In the latter, students will spend one year in each of two European institutions and return to Glasgow for year 5. Studio practice within the new programme is again assessed through judgements of work produced for a portfolio. Similar assessment criteria are in operation, although the assessments are reported and recorded somewhat differently.

There are three aspects of assessment in product design which are clearly different. Firstly, students take one European language in Year 1 and a second in year 2; the courses are taught by tutors from Glasgow University language departments; a dedicated programme is provided for these students at times convenient to them. Assessment is by the methods conventional in language learning: written translation exercises, oral presentations and so on. Secondly, students study social sciences and psychology, and again these are provided by Glasgow University staff. Arrangements for the delivery and assessment of these subjects are currently being developed from their original form; a central principle in the emerging scheme is that understanding of social/psychological aspects of design is best assessed **through** the design projects. (This has echoes of the view occasionally expressed by interviewees in other specialisms that historical and critical studies should be integrated into the studio programme rather than pursued in parallel. The disadvantage of this is that students never ‘escape’ from their chosen specialism to work alongside students from other courses and schools.) Thirdly, in years 3 and 4 of the MEDes student assessment will clearly be a matter for partnership with a range of European host centres; this requires not only complex administrative arrangements, but also that the meanings attached to the operative assessment criteria will have to become the subject of shared understanding. A common language of assessment will have to develop.

The challenges that can be involved in translating descriptive judgements of levels of performance into degree classification decisions have already been mentioned. These occur again in product design with the additional complication of the European language and cultural dimensions.

We tend not to mark projects i.e. give them a numerical grading. We create profiles of students' developing abilities ; I suppose this is a form of descriptive categorical assessment. Our external partners don't say things like 'First class' or '2.1 standard'; however, they do say 'excellent in such and such a respect'. So we use the language of excellent, proficient and so on and provide descriptors of what these words mean. We need to develop and use an assessment language which is intelligible and congenial to all our European partners.

Product Design Engineering

This is a distinctly different course, run jointly with the Mechanical Engineering Department of Glasgow University. The degrees awarded are BEng (in 4 years) and MEng (in 5 years). The programme and the qualification are accredited by the Institute of Mechanical Engineering and successful graduates acquire Chartered Engineer status. The time of students is divided between the main Gilmorehill Campus and the Glasgow School of Art. The programme is intensive and demanding and the effectiveness of student selection procedures

(which operate according to different requirements for entry from the rest of SODAC) has to be high to ensure admission of students likely to be able to cope with heavy demands.

The design studio programme (within the School of Art) occupies approximately 50% of students' time and is assessed in ways similar to those already described for other students of design. The engineering component of the programme is assessed as other engineering courses i.e. through written examinations, laboratory and project reports and so on. The assessment system was summarised as 'GU assessment based on exams; GSA assessment based on studio practice'.

What is learned within the engineering courses is expected to be applied within studio practice. The assessment criteria fall into four broad categories: Design (which embraces aesthetic and cultural aspects); Communication (covering a range of visual and verbal capabilities in presenting ideas and products); Engineering; Management (of resources, time and laboratory work).

(b) What is seen as good in the current assessment of students?

This section aims to capture what interviewees said in response to the question 'What in your view is good (i.e. pleases you) about how students of design are assessed?' The intention is to summarise the positive judgements made of assessment practices and procedures; it should be clear that the researcher was not acting as an evaluator, but as a recorder of other people's evaluations. Before embarking on this, three general points are necessary.

Firstly, most interviewees found the form of the question unexpected. Assessment was a thing people did, thought about, discussed and argued over; it was unusual to be asked to praise assessment arrangements. However, after initial hesitation and some pause for thought, people found no difficulty in identifying things they found pleasing; on average each person spoke positively about four features of assessment. Secondly the views expressed by interviewees were open, frank and dispassionate. It would have been understandable if people had produced some enthusiastic, in-group rhetoric designed to convince this 'outsider' researcher (and perhaps even themselves) that everything in the design garden was admirable. Only in two very brief episodes within the nine lengthy interviews did this interviewer feel he was there to be persuaded rather than enlightened. Thirdly, aspects of studio practice assessment may appear to be praised more frequently than aspects of historical and critical studies assessment; this was to be expected given the composition of the interview sample and must not be construed as any kind of judgement on the relative merits of the assessment procedures in the two components.

Most of the pleasing features could be placed within four broad categories. A category was formed when at least five people (i.e. a majority of the sample) provided opinions classifiable under that heading. The four headings are: regular dialogue with students about assessment; the value of the assessment framework; the avoidance of unfairness through collaboration; the degree show as a public event.

Assessment involved regular dialogue with students.

All but one of the interviewees identified as good that assessment was in one way or another related to regular interaction with students whom they got to know well. The following quotations give a sample of the views expressed.

We use assessment throughout the course as a prompt to dialogue. Assessment is a shared thing. After every formal assessment there are tutorials to discuss student

performance. We meet them all and talk things through. Assessment is a shared, transparent process that we're heavily involved in on a day-to-day basis. When we are talking with students about their project work, teaching and assessment are completely enmeshed.

We talk a lot with students about assessment; our assessment procedures are seen by students to be as fair as we can make them because we take pains to develop student understanding of our assessment system.

We are building up a relationship with students, detecting areas of strength and weakness and talking about these. We give them almost continual feedback. And we ask them to self-assess – to prepare an assessment form before coming to talk with us about their progress; and we discuss with them any significant differences of opinion.

We have to help students understand the language of assessment – and we do that through talking about our assessment criteria and prompting them to talk about their work in these terms. There has to be dialogue to get shared understanding of these criteria.

Immediately after this interview with you, I'm going to discuss with a student our latest assessment of her performance and progress....

This association between assessment and dialogue was very striking and did not occur in any of the other three subject areas investigated. Within this general heading, members of staff seemed to be saying four distinguishable things. Firstly, given relatively small numbers of students and the realities of studio practice, staff talk regularly and frequently with students; in the teaching and learning that goes on within a studio it is unavoidably the case that informal assessment (and self-assessment) occurs. During a short guided tour of one studio, the tutor-interviewee was asked by one student, 'What do you think of this...?'; the reply was 'It looks good to me. Perhaps you should be paying more attention to X? We'll talk about it this afternoon.' This can be seen as a commonplace student/tutor exchange, but essentially it is a rapid bit of diagnostic assessment, followed up by constructive feedback and the promise of further dialogue. Secondly, interviewees were saying that right from the beginning of the course, i.e. before any assessment occurred, tutors saw the need to help students understand both assessment procedures and assessment language and this was done not only through explanatory documents but through discussion. Thirdly, in the actual process of formal assessment (e.g. interim and final assessments) it was normal to have oral exchanges with students ('Tell me what you were trying to achieve in this? Why do you think A is better than B? Is this work better than you produced last year? In what ways?') Fourthly, after relatively formal assessments had occurred, it was normal practice for tutors to discuss with students the results of the assessment.

Underlying these comments is a distinctive view of what assessment itself actually means. This is obvious when the comments of design staff are contrasted with those from staff in one of the other subject areas studied. In that area staff saw teaching and assessment as clearly separable activities, with the latter detached from, and taking place **after**, the former. This view is easily held only when assessment is regarded as synonymous with testing and examinations. Design staff did not construe assessment in this way and thus it was easy for them to see assessment as occurring **during** teaching/learning. This way of thinking was vividly expressed by one person.

When teaching is thought of as delivering things, then examinations can test whether or not students remember their being delivered. I am very cynical, or perhaps just realistic, about examinations – but I am very positive about assessment.

Examinations were seen as things that staff did to students, whereas assessment was seen as something that was done with them. Assessment thus conceived required regular dialogue with students.

The assessment framework was valuable

The term 'assessment framework' is used here to cover the three (initially four) broad assessment criteria, the 16 or 18 sub-criteria and the five levels of performance, as described earlier. This assessment framework was explicitly praised by two-thirds of those interviewed.

The framework was seen as having 'evolved gently over time', this evolution being prompted by discussions within and across design specialisms and being guided by student reactions and feedback. The framework helped students to understand what was involved in the design process and to recognise that how they were being assessed was in harmony with the nature of design. The continuity of the framework over the duration of the course encouraged in students an awareness of both coherence and progression. The framework helped with comparability across specialisms and with Historical and Critical Studies. However, within the commonality there was room for specialist interpretation and discipline-based differences of emphasis. In a subject where subjectivity and personal preference were always going to be present, the existence of clear assessment criteria helped make explicit and overt those aspects of design in which some objectivity was possible. As the assessment framework had become more familiar and its implications better understood, it had helped to reduce (but not remove) time-consuming debate amongst staff and between staff and students; it had, for example, helped to structure feedback and thus make its provision more economical of staff time.

All this builds up to a remarkably positive overall judgement. However, it was clear within the data that the current position had been reached after extensive periods of reflection and debate, which had been punctuated at times with intensely emotional episodes. At times, one or other specialism had been seen either as 'pushy' or 'intransigent'. Clearly also the evolutionary change was continuing and there would be further skirmishes or battles. The extent to which further development might be in response to external influences is discussed further in Section F. Illustrations of typical views on the assessment framework are as follows.

It is helpful to students to analyse the design process like this. In any subjective subject like design, there will always be difficulties in describing the assessment criteria. It helps to be prescriptive about where it is appropriate to be objective. It was pretty clear from students that they didn't like this format at first, but they got used to it and now see it as useful.

We have thought very carefully about assessment and adjusted it in response to student worries. We have worked hard on our understanding of the pro-forma and the criteria and on student understanding of how such a framework can be formative for their own development. Students are now pretty clear as to what is expected of them.

The fact that people found the assessment grid helpful did not prevent them from suggesting ways in which it could be improved. These usually focused on the addition or deletion of one or two sub-criteria or subtle changes in wording. There was one variation on the grid that this researcher found particularly fascinating. There had been a recognition that the usual forms of the grid (using ticks in boxes and bar-chart diagrams) tended to encourage a static, snap-shot view of student abilities; thus there had been developed the use of 'polar charts'. In these circular representations the criteria and sub-criteria were featured round the circumference

and development levels were represented by radial distance from the centre. Assessors then indicated their judgements of level of performance on each relevant criterion, allowing the construction of a simple join-the-dots profile. This apparently simple modification made much simpler the identification of unevenness in the profile of developing strengths; growth in student abilities over time was easily displayed, as was inter-assessor variability.

Unfairness is mitigated by collaboration

Five of the nine interviewees identified as pleasing that their assessment system managed to lessen the chance of unfairness to individual students. Why then did they think that unfairness was a potential hazard? Students worked closely with a small number of tutors; the role of the tutor was to enhance students' learning and development – in this process informal assessment and constructive feedback were key features. Then at two significant points in the year, the tutor acquired a more formal assessment responsibility. It was recognised that in this type of assessment there was a possibility of assessor bias operating either to the student's advantage or disadvantage. (It was notable that in one of the smaller departments arrangements were made for students to change tutors at the end of term 2, so that the 'lead' tutor for the majority of the year was not the lead tutor during the final assessment.) The nature of design was also seen as relevant here. It was inevitable that all individual students and staff would have different tastes and personal preferences; these could enter the assessment process as sources either of positive bias ('I am rewarding this because I personally like it.') or negative bias ('I am aware that I like this; I must make it obvious that I am not over-rewarding it.')

The possibility of assessing students unfairly was avoided, it was said, by always having a number of assessors, especially for the Final Assessment.

Recently I counted up the number of people who would look at the student's work before the final studio grade was awarded. It came to I think 23! It went like this. Those with in-depth knowledge of the specialism, that's four in our case, assessed each student's portfolio independently according to the criteria. Then they had a discussion over coffee and came to the first internal grading on each criterion for every student. Then all the heads of department looked at all the other students in the other departments; they could suggest that any assessment be re-visited by the specialists. Then the specialist external examiner marked each student's work 'blind', then the internal and external assessments were put together, final say going to the external. Then the group of external examiners toured round the work of all students in all specialisms. And then after that there's the Board where the studio mark and the H and C mark are put together to reach a degree classification.

This account may suggest the process is a somewhat mechanical business that routinely delivers an 'accurate' verdict. The reality may well be rather different and there is no doubt that this collaborative process would provide fertile territory for observational research. Some hints about the character of the exercise can be gained from the quotation which follows.

When we've done the internal assessments as a department, the heads of department walk round as a group and that's really quite interesting. They are supposed to be confirming our judgements about the good ones, the poor ones and the ones on the cusp, before the External Examiners are involved. This is a delicate matter. You have to articulate the reasons for your internal marking to those from outside your own area; this is to encourage commonality and comparability of standards. The conversations are not really objective and analytical; people tend to say 'I'm a bit surprised you're only giving that a third' or 'Is that really as good as you seem to think it is?' It's a fantastically sensitive time of the year. We're a bit lucky here in

that people don't really understand what they are looking at – so we can get away with it!

Unfairness is, of course, likely to flourish in dark and secretive places; the extensive opening-up of the assessment process, giving many people the opportunity to collaborate, scrutinise and discuss, undoubtedly makes unfairness less likely. There were perhaps some over-easy assumptions amongst some of those interviewed (a) that consensus was the same as 'accuracy' and (b) that judgements were made in a wholly rational way in which observable evidence played the significant role and not the relative power of individual assessors. There was also the suggestion from two people that the greater the number of staff who were involved in the assessment decision making the less likely it was that students would dare to question the final outcome.

The assessment process culminates in a public event.

Five of the nine interviewees identified as 'good' that fact that the assessment process culminated in a significant event, 'The Degree Show', open to press and public. It was emphasised that this was a 'real' exhibition; it was not just that students had to present their work to the examiners 'as if' it were for public display. The School of Art degree shows have been an important feature of artistic and cultural life in Glasgow for many decades and they attract large, lively and vocal audiences of current and former staff, current and former students, press critics, talent spotters, potential purchasers and so forth. It was also emphasised that the audience attracted was not a local, parochial one, but a national, European and American one.

'The Degree Show' was being talked of as an element of the assessment process, interviewees saw it as such. It should be asked why this should be so, because one could argue that student assessment finishes at the conclusion of the Board of Examiners' meeting where degree awards and classification are decided, and this is a few days before the degree shows. The most frequent comment from those interviewed was that the existence of the degree shows motivated students to produce work of a high standard; they were part of the mechanism for ensuring that high standards were maintained, for demonstrating to a wider public that standards were being maintained and for confirming that staff were able to stimulate and guide students to produce good work.

For our students, going public is a naked sort of experience, but the thought of it does motivate them to a high level of performance. There's also a sense in which it demonstrates that we are doing our job. I say to the sceptics: 'Come to our degree exhibition. If you don't find quality there, I shall resign.'

It was noted above that in the final assessment of students the number of assessors involved increased remarkably; both staff and students saw the degree show as an extension of this process – the public at large have some voice in the overall assessment of students. Staff spoke of this as being related not only to accountability but also to democracy: assessment was legitimately moving from the closed and professional to the open and public. Of course professional and specialist judgement was not always confirmed and validated. When external, critical comment did not corroborate specialist judgement, staff had a very convenient riposte: 'we have more evidence on which to base our overall verdicts on students than is made generally available'. Less cynically, staff saw public comment as an additional influence to be taken into account in the unending task of refining the meanings to be attached to their assessment criteria. Just what is 'good' design?

Intriguingly a couple of people described the degree show as an important formative influence on students. The degree show cannot be seen as a means of formative assessment in any conventional sense; however, as noted earlier, it can be regarded as formative in that students do learn from it in ways that help them in their future careers.

Assessment culminating in this kind of public event has no exact parallel in the other subject areas studied. It would be unusual, for example, to think of graduation ceremonies as a part of the assessment process; rather they are a formal celebration that once again assessment has been completed and that certification of attainment can properly be issued. The kinds of message being publicly transmitted are, however, rather different. At a graduation the message is from the institution: 'This university is giving the students something....'. At a degree show the message is essentially from the student: 'This is my work – you too may judge it'.

There were other things, each identified by a minority of interviewees as 'good'. These were:

- the choice of weighting for Historical and Critical Studies in the final year;
- the welcome emphasis on 'exit velocity';
- the authenticity and predictive validity of portfolio assessment;
- the value of subject-specific external examiners;
- numerical and formulaic aggregation of assessments was largely avoided.

There was approval for the relative importance attached to the assessment of studio practice and historical/critical studies in years one to three and for the provision of some flexibility in year 4. That students could elect (after guidance) to provide either the Extended Essay (15% of total) or the Dissertation (30% of total) was seen as good.

It seems to me we have a pleasing balance between the creative and the theoretical – and I do like the option for some people to choose to give more or less weight to written work in year 4. A large proportion of our students do find essay writing a struggle – dyslexia is a very real problem; they tend to go for the 15% option and thus get it out of the way earlier.

It's not just 'you're good at essays – take the 30% route'. Students react to H and C very differently. Some don't take it very seriously and do only the necessary minimum. Others find H and C a genuinely liberating alternative to studio practice; they want to do more – and they can be superb.

Staff liked the significance attached to 'exit velocity'; this didn't mean they liked the term. It was seen as a pleasing feature of assessment within the degree programmes that students could make mistakes, go astray at times without this having any influence on their final assessment (as long as they satisfied the progress requirements). The only thing that mattered each year, and in the programme as a whole, was the standards achieved in the Final Assessment. It was said to be common for students to get something of a shock on an Interim Assessment and then 'really pull their socks up' and achieve a standard at the final that was unexpected. This approval for exit velocity (i.e. you are only as good as your last assessment) was usually related to pleasure that the course was a non-modular one. Several interviewees with experience of modular courses elsewhere were highly critical of the fact that marks and grades obtained early in the programme ('before the penny dropped') could have an impact on final assessments.

In two of the other subject areas studied, some staff expressed a clear approval for continuous summative assessment; they saw it as preferable to big-bang final examinations. In design, where there was no question of terminal big-bang examinations, continuous summative

assessment came in for significant criticism. 'Exit velocity' within design meant the level of performance achieved gradually over time and reached **by-the-day** that the portfolio is submitted; in other areas 'exit velocity' referred to what students demonstrated **on-the-day** of their last examination. The two are very different.

Our process has a welcome fluidity; within it people can experiment and make mistakes. Sometimes the most worrying of our students at one stage come good at a later stage. So we are assessing them at the end of the process, when they are ready to leave us – we are giving them the chance of growing, of being educated, for as long as possible.

Giving them marks throughout a modular course can mean for some an early collapse of motivation. We recognise development right through to the end and that their exit velocity is what matters. Students must be allowed time to mature.

Portfolio assessment was seen as appropriate because of its authenticity; the criteria by which the students' work was assessed could be related to a generally acceptable analysis of the design process. Assessment within the design course was similar to the ways in which future employers or exhibition organisers would assess these students when they became qualified designers. It was claimed that the assessment made of students as undergraduates tended to identify effectively the level of success of the students in later life i.e. the course assessments had predictive validity. There was an element of self-fulfilling prophecy in this; students went on to succeed because the School of Art had labelled them as good. But the continually updated portfolio was a better indicator of quality than an academic judgement with its built-in obsolescence.

We do keep informed about the progress of our products. And our assessments are a very good indicator of how well they will do in their careers. It is intriguing how some of them go on to construct careers to suit not just their abilities but their personalities. And perhaps they have the confidence to do this because we have told them they are good..... But the portfolio is really the thing that tells people about their quality.

Subject-specific external examiners were seen as contributing positively to the general credibility of assessments. External examiners with a different area of design specialism might well have useful comments to make on assessment procedures overall or contribute helpfully to debate on the comparability of standards, but they did not have the experience, the legitimacy and perceived authority as same-specialism assessors. It was good for students to discuss their work with internationally acknowledged experts in their own field. Any move to generic externals on economy grounds would be seen as seriously retrograde.

The general pleasure that the Glasgow design course remained non-modular has already been mentioned. Two people said that modular course organisation resulted in assessment aggregation and decision-making that was 'numerical, crudely quantitative and formulaic'. That such approaches had been avoided was very pleasing. There was of course a measure of formulaic handling of numbers within Boards of Examiners when studio and historical/critical grades were combined for the pass/fail decision at the end of years 1,2 3 and when degree classifications were determined at the end of the final year.

C STUDENT ASSESSMENT: LOOKING BACK

Interviewees were asked to identify any changes that had occurred in assessment practice within the design curriculum; if they identified any one change as particularly significant they were then asked why, in their view, such change had occurred. It was anticipated that the latter would be a difficult question to answer and so it proved. Change in educational practice usually has a multiplicity of causes and these can be subtle and interacting. It is possible for people to be involved in change and not to know, or indeed much care, why it is happening; it is also possible that one knew at the time but cannot now recall the reasons with clarity. It is also likely that one's past perceptions of what the powerful factors were will have been modified with time. However, no apology is necessary for asking the question: people's perceptions of reasons are arguably as important as the 'real' reasons, even if it could be agreed that these are ever accurately knowable.

It was the general view of those asked that there had been little radical change in assessment practice during their time as assessors of design students at GSA: 'Well – I don't think so'; 'No'; 'Not really'.

Things haven't really changed fundamentally in the years I have been here – and I am not suggesting that they should have done.

There may not have been radical or fundamental change, but all interviewees did identify some changes that had occurred. The most frequently mentioned changes were in the assessment criteria applied.

There have been changes in the wording of the criteria, the meanings we attach to the words and in the weightings of individual criteria.

The criteria have been honed in an evolutionary sort of way and so too have the ways they are grouped and represented on the assessment report forms. We keep tweaking the design of the assessment pro-forma.

Five other changes were identified more than once. Firstly, over the years increasing attention had been paid to the development and assessment of oral and written communication skills. Secondly, there had been a growth in the awareness of staff of the importance to student learning of student self-assessment. Thirdly, there had been the welcome introduction of a choice in the weighting (either 15% or 30%) of the written work on Historical and Critical Studies in the final year. Fourthly, there had been a growth in insistence within procedures that assessment decisions should be better evidenced; assessment practice not only had to be fair, but it had to be possible to demonstrate its fairness if that was challenged. Fifthly, the documentation surrounding all aspects of assessment practice had been improved. (Although one person thought the documentation was not better, merely more abundant.)

There were three explanations offered for these changes: there had been some 'gentle pressure' to encourage the sharing of good practice within the design school; listening to students and being receptive to feedback from them had resulted in some changes in the criteria; responding to pressure from the Quality Assurance Agency had brought about changes in the language and documentation of assessment, but not in central aspects of assessment philosophy or procedures.

It was illuminating that one person drew a clear distinction between assessment practice in GSA and his/her own previous experience of assessment. S/he had come fairly recently from a college with a modular form of course organisation in which every student activity and project in every module required an assessment report; this, combined, with a large rise in student numbers, had made assessment a 'nightmare' not only in the marking and recording

of large numbers of separate assessments, but also in the soul-less and impersonal deployment of mathematical, formulaic methods of aggregation and decision-making.

There was never time to talk to students; assessment is so futile when it doesn't take place within a context of relationships and dialogue. Here it was a great relief to find much more depth in how students were assessed. I would be totally opposed to modularisation of courses in Glasgow. An external assessor from RCA said that Glasgow was strong in assessment because it had retained one-to-one relationships within a non-modular course-organisation.

There were two areas in which there had been very significant changes in some aspects of assessment: product design and product design engineering. The introduction of new degrees (BDes, MEdes, BEng and MEng) had inevitably required a general re-thinking of assessment procedures and arrangements. The new degree courses were attempting quite different things from the ones they replaced; they crucially involved working in partnership with other institutions, in Glasgow and in Continental Europe, which had widely different assessment traditions and cultures. Even here, however, the ways in which the design studio practice elements of the courses were assessed had probably not changed dramatically. True, there had been a number of changes in the main assessment categories and their constituent criteria and in the ways these were represented. Product design and product design engineering had 'departed somewhat' from the assessment grid used elsewhere in the School of Design. Nevertheless it seemed to this writer that in its main features and emphases the assessment system for design practice remained broadly similar, depending as it did on portfolio assessment, on a matrix of criteria and levels and on dialogue with students. There would be further changes in assessment: for example, the assessment of how well engineering knowledge was **applied** needed to be enhanced, the assessment of language capability and communicative competence could be improved and there was some need for a more clearly conceptualised distinction between the analytical, the creative and the development parts of the design process. It might be that the changes made and anticipated in product design and product design engineering would become useful catalysts for further developments in the other five areas of design. Although the two areas had 'departed somewhat' from the others, they had not separated off and severed lines of communication.

The first quotation in this section (on page 19) made a strong impact on the researcher. Perhaps assessment within design **should** have changed more than it has. This is discussed further in Section E.

D STUDENT ASSESSMENT; LOOKING FORWARD

(a) Changes that people wished

Those interviewed were asked what changes they personally would like to see in how students were assessed, assuming that the conditions for introducing such changes could be made favourable. With only a few exceptions people had to think hard to come up with suggestions about **any** change in assessment they were keen to see. In the majority of answers there was a pervading tone of satisfaction with existing systems and procedures: 'We're really quite happy with assessment the way it is'; 'I can't really think of any change that seems at all urgent'. This sense of satisfaction was with a process which, although recognised to be difficult and time-consuming, was well-done in ways which met with broad approval; justified satisfaction should not be confused with unthinking complacency.

We choose to allocate a lot of time and resources to the process of assessment; we do this because the importance of assessment is recognised, not just by me but by all the staff involved. We do talk and think a lot about assessment as well as actually doing it.

It was notable how many people made some reference to time in their replies. Given more time, people would not do anything very different i.e. they were seeking a change in conditions rather than any radical change in assessment practice.

I'd like more time to help students at the beginning of Year 2 to understand the how, what and why of how we assess them.

I want more time between assessment and the provision of feedback to reflect on what it will be best for me to say to students.

We need more time for project assessment during the year – we can all make the time for the final assessments.

Some students need more feedback, but others need more input; perhaps that is teaching, but I can't separate teaching from informal assessment.

There were only five suggestions for any definite change in assessment practice and each of these came from only one person. Firstly, there should be greater comparability of levels and standards across the five specialisms of the BA (Hons) Degree; this could be brought about, it was said, by bringing all the External Examiners together earlier in the assessment process. As it is they enter the process too late to have much influence over the internal assessments. Secondly, action was required to reverse what one person perceived as a drift in standards; too many people now do honours degrees and it is too easy for people doing honours to achieve first and upper-second class degrees. Thirdly, in Historical and Critical Studies there should be more scope for student production of non-standard materials for assessment; this would require the existence of negotiable outcomes and considerable ingenuity in the definition and operation of assessment criteria; it would be difficult to make this acceptable to QAA who were over-concerned with standardisation and comparability. Fourthly there was a plea for better methods for the assessment of new subject material relevant to the context of design; there must be more satisfactory ways of assessing it than treating it as detached and separate subject content. This applied particularly to languages and social subjects within the newer product design degrees. Effective integrative assessment methods could perhaps be devised, but to make these of high quality took considerable time, expertise and goodwill amongst the partners involved.

The fifth suggestion for change was made in the context of product design engineering, but it could probably apply to all areas (not just of design but of higher education). The suggestion deserves to be quoted at length.

We all manage to get through our assessment procedures because of our familiarity with what we need to do. We have all these bits of paper; they are all, about the same thing essentially, but they should all be speaking to one another rather better. There ought to be consistency of language in the calendar, in the briefing documents, in the assessment sheets and in student records. I want to make sure everything is coherent, with logical progression for students from one stage to the next. And not just coherent, but explicit and conveying clearly to students a clear sense of the standards required at each stage. It's all there somewhere or other, but are all the relationships clear? Can we put it up on the wall? There it is, folks - all you need to know about 'assessment for progression'.

(b) Changes that may be required.

There is a possible ambiguity in this sub-heading. This section is not about change that is 'required' in the sense that it is needed if an existing situation is to be improved; it is about changes that may be **required of** design staff by some authority external to them. The previous section dealt with answers to the question of what assessment change people would personally like to see; for this section, the prompting question was 'What changes in assessment do you think that you may be required to make, whether you like them or not?'

The general answer to this question was quite clear.

No – there's no big change likely to be required of us. We have adjusted sensibly to demands in the past. I don't anticipate or fear any major pressure or upheaval...

There were only two changes which might be demanded of people that were given more than a single mention and even so, these were only mentioned twice each. The first was that an undesirable degree of conformity of practice across the separate design departments might be imposed.

The new directorate of GSA has an obsession about everyone doing assessment in the same way. They haven't got their minds round the fact that each department has a different focus. We're all design – but we are all delivering different kinds of curriculum, looking at very different things.

I rather fear the pressure towards commonality. We are developing the assessment criteria, our procedures and our ways of reporting and they are becoming more and more appropriate for us. It would be a pity if the School of Design forced uniformity across departments simply in the interest of common ways of reporting and of satisfying quality assurance.

The second required change might be towards the use of more numerical and quantitative assessment to allow more systematic decision-making. This fear seemed to stem from knowledge or experience of modular curricula and a general distrust of any assessment involving numbers ('the wholly spurious validity which some people attach to meaningless numbers'). It should be noted, however, that a dislike of percentage marking was not quite universal: 'I like percentage marks; there are so many of them, it gives me scope for making fine distinctions between people'.

All the other changes identified occurred as single mentions: data protection requirements would mean even greater care about what it was wise to write down; there might be an attempt to save money by moving from specialist external examiners to fewer generalists; current movements towards a more accessible curriculum would require a re-think of the place of assessments of written work, both in selection of students for entry and during the course; there might be a requirement for more assessment of oral presentations; the inclusion of too many live-projects (which were income-generating but not always of great educational value) could skew the assessment process; all changes in assessment tended to result in more paper-work – the increasing bureaucracy of assessment might have to be resisted.

The sources of pressure for change were seen as being the School of Design, central management in the Glasgow School of Art, the Quality Assurance Agency and what might vaguely be called 'the government'. There was no detectable emphasis on one of these rather than the others.

E INTERPOLATION: THE DESIRABILITY OF CHANGE

Earlier, attention was drawn to the quote: 'Things haven't really changed fundamentally in the years I have been here – and I am not suggesting they should have done'. In the previous section it became clear that there was no strong general feeling that any significant change was desired or likely to be imposed. In addition there was a recognition that assessment change would have been necessary if pressure for the modularisation of courses had not been successfully resisted. Is it possible that those responsible for assessment in design courses have somehow managed to exist in a backwater, seeing no need to change the practices of decades and unaware of the external influences that make change if not inevitable, then highly desirable? It would be easy, and quite wrong, to represent the situation in this way. There are general, theoretical reasons for this; but importantly the reasons are specific and evidential, grounded in the data provided by the design staff.

The idea of change in assessment practice is central to the present research; it was always likely that some interviewees would think that the researcher believed change to be 'a good thing'. There were people in all four subject areas who seemed to be expecting the interviewer to judge them on the number and nature of the changes they had, or had not, witnessed or instigated. Despite reassurances in the introductory letter, in the protocol for engagement and in the preamble to the interview, these people revealed some degree of uneasiness about the researcher's intentions. A few remarks were apologetic ('I'm sorry, we don't seem to have changed much'); some were defensive ('There has been no time to introduce the changes we may have wanted'); some were jocular ('Well, have I passed?') and a very few were supplicatory ('We need someone from outside to tell us how to do better').

It is now commonplace for people to operate on a set of linked, but largely unexamined, assumptions about change: change is essential; all changes are desirable; the absence of change is reprehensible. Overt indicators of these assumptions lie in all-too-familiar clichés: change is the only constant; to stand still is to go backwards. The assumptions exist not only in vox-pop soundbites. A leaflet recently landed on this writer's desk advertising a professional development course on assessment for university lecturers. This opened with the assertion that 'Clearly the assessment methods that have served us well in the past are no longer sufficiently robust in the current and future climate'. This could of course be well-founded, but it does deserve careful scrutiny. Is it so clear that there was some golden age in the past when assessment methods served us well? Is robustness the only or most important criterion for judging assessment methods? What is it in the current climate that makes formerly effective methods no longer satisfactory? Can predictions about the future climate be right overall, but wrong for the micro-climates in individual subject areas?

The relationships between judgements about change and judgements about quality at specific times deserve to be unpacked a little. For simplicity, accept the period of interest as being say the ten years from 1990 to 2000. Suppose the simplest binary decisions between 'good enough in the context' and 'not good enough in the context' were made in 1990 and again in 2000. There are then four possibilities for a summary longitudinal evaluation of assessment: progression/improvement, regression/deterioration, maintenance at a satisfactory level and maintenance at an unsatisfactory level. Suppose now a crude, global judgement is made about whether the context of assessment has changed significantly or not; relevant features of context would include changing purposes in teaching the subject, the number and nature of students, the number, expertise and priorities of staff, the departmental culture and the institutional climate, institutional and national policy requirements, the availability of resources both human and technological and the expectations of public accountability. Next consider judgements about whether or not there have been significant changes in assessment practice over the period and then judgements about whether or not the changes have been effective ones. Accepting all the simplistic dichotomies and heroic assumptions in the above,

there are twelve main 'histories' a department could have with regard to its assessment practice. (Given the four parameters one might have expected sixteen, but if there has been no significant change there is no need for consideration of whether the change was effective or not.)

At the risk of labouring the obvious, two sample histories of fictitious departments where there have been no significant changes in assessment should be contrasted.

In department A assessment practice in 1990 was good enough
 there has been significant contextual change
 there has been no significant assessment change
 assessment practice is not now good enough.

In department B assessment practice in 1990 was good enough
 there has been no significant contextual change
 there has been no significant assessment change
 assessment practice is still good enough.

This simple illustration makes the point, perhaps a banal one but one too frequently overlooked, that although the process of change is a common focus of interest, the satisfactoriness of a present state can not be judged by whether or not recent change has occurred to reach that state. It is possible for assessment to be in a satisfactory state without recent change having occurred; conversely assessment may currently be unsatisfactory even although assessment has changed significantly.

The judgement of external evaluators, for example during teaching quality assessment and subject review, might well be partly determined by the view that any claim that there has been no significant change in the context of assessment in the last ten years is simply unsustainable; if that view is accepted it is of course more difficult then to defend the absence of any significant change in practice. They might also argue that there is no situation so good that it cannot be further improved. Two possible lines of response spring to mind: (a) the situation is not as excellent as it used to be, but still better than most and (b) the standards for judging adequacy in assessment should be relaxed in recognition of the increasingly difficult context.

The perspective of the design staff interviewed here can be summarised as follows:

- (a) our assessment procedures were very good in the past;
- (b) there has been little change in the context of our operations that impinges significantly on assessment;
- (c) there has been little significant change in our assessment practice;
- (d) our assessment procedures are still good;
- (e) there is no pressing need for much future change in our assessment practice;
- (f) given more time for assessment we could do what we currently do, but rather better and also make those minor evolutionary changes that we recognise as desirable.

Three quotations illustrate most of the features of this general summary.

Some may think we operate an old-fashioned system of teaching and assessment based on a 1:1 relationship with students. We still have approximately the same number of students, albeit with fewer staff, so that has not been a big pressure on us to change. We've got an assessment procedure which works pretty well because of its validity, so we've retained it. We've been able to do that because the course has not been modularised and GSA is popular because it hasn't modularised. There would only be a need for radical change in assessment if we had any big change in degree

or course structure. The rest of the world has shifted, but I don't think it is any better for having shifted. But sometimes one feels that if one hasn't done something new recently one can be judged to be slacking.

We have always assessed in the ways towards which assessment was evolving elsewhere; the key word is authentic....

We can find time to do the assessment; there isn't much time to spare for thinking how perhaps it should change and develop.

F ASSESSMENT: FACTORS INFLUENCING CHANGE

In this subject area it was not possible to focus on one or two significant changes and explore with people why these had occurred. In so far as interviewees identified causes of change and factors influencing change, they emphasised three: sharing of good practice and learning from others within the School of Design; the policies (or perhaps the perceived enthusiasms) of senior management within the Glasgow School of Art; the requirements of the Quality Assurance Agency (and earlier bodies like the Higher Education Quality Council and the Council for National Academic Awards). The situation was rather different with regard to product design and product design engineering; here assessment procedures had inevitably been re-considered as part of the whole process of course planning for an innovative degree. The degree as a whole was different; the assessment system was different, especially in relation to components other than studio practice.

After trying to elicit people's own views on factors influencing assessment change, the interviewer presented each person with a list of factors identified in the literature and asked for a reaction to these. Interviewees were asked (a) to give their view of the importance of each factor in relation to assessment change in design in the past and (b) to predict which factors would become more influential in the future. One of the nine interviewees felt unable to respond; the results for the other eight are summarised below.

Factors influencing past change in assessment

Factors for which the most frequent response was 'very important in design'

- Concern for the quality of student learning
- One enthusiast persuaded us
- A general feeling that change was desirable

Factors for which the most frequent response was 'of some importance' (and on which there was a spread of opinion)

- Increased diversity within the student population
- Increased number of students
- Awareness of trends in assessment elsewhere
- Educational/assessment theory or theorist
- A change in course structure

Factors for which the most frequent response was 'not important in design'

- National policy required change
- Emphasis on transferable skills

- Faculty (i.e. School of Design) policy required change
- Institutional (i.e. GSA) policy required change
- Pressure from external examiners
- Students pressed for change
- Policy from an external professional body
- The availability of appropriate assessment technology

Factors predicted to become more influential in future

- Increase in student numbers
- Pressure of student opinion
- National policy ('the government' and QAA)
- Institutional policy
- Availability of appropriate assessment technology

By this time in the interview, everyone was used to giving elaborated and thoughtful responses; when the interview moved into this more structured phase the majority tended to expand on their answers and not just provide tick-the-box type replies. These elaborations allow the following comments to be made.

The three factors most frequently cited as very important make an interesting group of influences. Concern for the quality of student learning was said to be 'what drives everything we do in teaching and assessment' and 'to be the key reason for any change in our practice'. A majority of people said that, importantly, enthusiasts in the design school had persuaded them. The picture seemed to be of three or four people, more interested in assessment than the others, who had no great difficulty in encouraging those others to consider and embrace change (perhaps because the changes were quite modest in scale); the enthusiasts were not seen as passionate champions blazing some highly original trail. Things had happened because there was a general feeling that some modifications were desirable. The interview data provided a pervasive sense of people doing what they did because they themselves thought it appropriate; the change factors were largely internal.

It is perhaps surprising that concern for student learning was rated as very important and student pressure for change as not important at all. Staff made it clear that they listened carefully to what students said; they took it seriously but did not regard it as providing much pressure for change. (One person did say that the student voice had been important in 'the tweaking of the assessment pro-forma'.) There are different ways in which this can be interpreted and further evidence would be necessary before commenting on their relative credibility. Perhaps staff have a view of student-centredness which doesn't extend into debating the nature of assessment with students; perhaps students don't say much about assessment – or more probably say plenty about the results of assessment as they personally affect them but little about the procedures by which results are generated; perhaps students are very satisfied with assessment procedures and have no good reasons for seeking change.

Intriguingly some staff spoke of external examiners in very similar terms: 'On the whole we teach them about assessment rather than the reverse'. The interview data do seem to provide a picture of people with considerable confidence in what they do; there is no evidence in the data that the confidence is other than well-founded.

Staff have not become isolated and immune from external influence. Such influence was particularly obvious in the areas of product design and product design engineering; the design, planning and introduction of new degrees had required considerable attention to thinking and practice elsewhere (and not just within the UK) and study of the relevant

literature in professionalism and in assessment. In some of the other design specialisms knowledge of assessment issues in England and USA had been influential.

It was noticeable that policy influences, local, institutional and national were said not to have been important influences on assessment change. (One important exception quoted was the influence of CNAA policy at the time of re-validation of the BA degree.) In contrast with the subject areas of chemistry and medicine, there was no single professional body in the design world which had an influence on assessment practice; the nearest example here was that of the Institute of Mechanical Engineering on product design engineering.

It was predicted that five influences on assessment would become more important in the future. These were institutional policy (from the central directorate of GSA) and national policy (vaguely from 'the government' and more specifically from QAA or whatever body might replace it). Students would have more impact both if their numbers relative to staff continued to increase and also if they became more vocal in their opinions on assessment; the latter was recognised as a general feature in higher education ('students everywhere are becoming more demanding and even litigious').

G. EMERGING ISSUES

In this section two 'big issues' are identified and discussed. They are expressed initially as questions:

How can creativity in design be validly assessed?

Can classified honours degrees in design be justified?

The researcher had no pre-conceived ideas that these would emerge as important issues; questions about them were not included in the interview schedule. These issues emerged as topics which were of importance to the people interviewed; they talked about them with enthusiasm recognising that their opinions might well not be shared by others. It should, however, be noted that the first of these issues was of particular interest to the interviewer and his interest may well have prompted more attention to it than otherwise it would have received. As one person said of assessment itself: 'Like a lot of things, the more you talk about it, the more interesting it becomes'.

(a) How can creativity in design be validly assessed?

This is expressed as one single question, but it could well be deconstructed into a range of subordinate questions. What is creativity in design? How important is it within design courses? How is its existence recognised? Will different assessors agree when it occurs? Can students be protected from possible injustice in the assessment of creativity? How can local judgements of creativity acquire some general validity? A comprehensive and satisfying answer to all these questions would need an extended exercise of research and scholarship well beyond the scope of the present study. What will be attempted here is some organisation of the views of interviewees in a way which makes a starter contribution to the provision of answers.

It has to be said in preamble that there are no simple answers to these questions which would be agreed by all the interviewees.

Everyone knows what engineers are; no one knows what designers are. They wear coloured shirts and they wave their arms about a lot – beyond that there's disagreement about everything.

It is difficult to distinguish the genuinely creative from the off-the-wall – that's what we regularly argue about.

An essential element of the creative in design was said to be that the student **generated** something of their own; being creative was a matter of production rather than reproduction. There could be a problem in identifying what was merely re-produced; there were so many high-style cutting-edge design magazines now available that students could 'borrow' from them without the source being recognised by the tutor or assessor. Such borrowing did not amount to plagiarism unless there was dishonesty and the intention to deceive; it was after all a feature of almost all design to remember something, and then to borrow it and adapt it appropriately to a new purpose or context. There is little that is wholly new.

Students of design have to generate something of their own. The situation is, however, rather different from that in say Fine Art. In design, the thing created has to meet a complex set of practical demands stemming more or less directly from the initial design brief; the judgement then required is not merely of the design's newness or originality but of its adequacy in meeting what is required of it. Those interviewed were quite clear that creativity was not about some courageous leap out of ignorance into the dark; rather it involved the synthesis of what was already known into something new and appropriate to its purpose. Such synthesis, although at base it remained personal and imaginative, was part of a recognised design process which had its own clearly explicated conventions in which students were expected to develop competence.

The original brief on paper evolves into ideas, models, sketches, preliminary designs and products. One of the most interesting things is seeing how students cope with the demands, creating some kind of order out of the very diverse bits of information they possess and acquire. Students with design acumen can put something together which is greater than the sum of its parts. This is not being creative in any 'arty' sense; it's what I'd call having the competence of creative synthesis.

Creativity might be a necessary and important component of design capability, but it was by no means a sufficient one. One of the broad assessment criteria had to do with originality/innovation/creativity, but there were many others and the possession of startling originality was not enough to compensate for the absence of other qualities

Sometimes a particular project may call for brave answers. But we have also to remember that it is not the most heroic designer who is the best designer. For example, people skills to do with relating to others and negotiating with them are sometimes much more obviously necessary. There is little point in a highly creative designer who cannot relate either to clients or to technicians.

How then is creativity recognised? It intrigues this writer that creative synthesis was labelled by one person as a competence. There may be an implication here that by calling the creative act a competence, its recognition can be made to appear less problematic than it actually is. Undoubtedly there are within design many skills and processes which are relatively easily recognised and assessed; there are right and wrong ways of doing some things, there are ways of doing others which are generally agreed to be more or less effective. Simple recognition and relatively objective assessment of these is possible. The competence of creative synthesis is, however, very different. One very obvious argument supporting this is the general acceptance that Category V behaviours within the Bloom taxonomy (i.e. 'synthesis') are not amenable to any kind of objective assessment.

It was possible for the products of design at first to appear wayward or bizarre. They could however be assessed as genuinely creative if the student designer were able to convince the assessors that this should be the case.

Is the design solution creative or anarchic? That's where the background material comes into play. And whether or not the students can argue their case logically and rationally. Plenty of advances in art and design have played on anarchy – but with a well-argued case.

The answer provided by interviewees was thus that creative ability was never unequivocally demonstrated in the products of design. The existence of creativity was inferred from the detailed knowledge of the processes by which the student generated the product i.e. the assessment of the creative element was possible because of the close relationship between student and assessor built up as the student works through the processes. This is convincing, but it does mean that the tutors and internal assessors base their assessments on a wide range of evidence, not all of which can be made available to others. When assessment widens out beyond the closely involved specialists, it is highly likely there will be tensions between those assessors whose power comes from the knowledge of detailed involvement and those whose power comes from their disinterested or institutional status.

(b) Are classified honours degrees appropriate in design?

The majority of those interviewed talked of their misgivings about the appropriateness of classified honours as the exit qualification from a degree course in design. There were four main objections to the current provision: class of degree was irrelevant in the wider design world; degree classification was seriously divisive amongst students; determination of the degree class was unreasonably time-consuming for staff; the absence of other exit points led to a devaluation of the honours awards.

Degree class was of no interest to future employers and clients.

In design no-one will ever ask you what class of degree you got, it is professionally neither here nor there. All future assessments of your quality are done via your portfolio of work and no-one gives a damn about degree class.

We are the only people amongst all our European partners who break down our degree results into classified honours. Outside people never look at someone's portfolio and say that's 2.2 standard work – they say that is excellent in that regard and satisfactory in others.

Degree class was not even a crucial determinant in decisions about eligibility to pursue other qualifications.

They say that the distinction between 2.1 and 2.2 matters for entry to teacher training – but does it? They take 2.2s who are promising in other ways. And the director of postgraduate studies at St Martin's College says she is only interested in her own judgement of an applicant's work; she doesn't give a damn if someone has a third. I myself would love to have just fail/pass/distinction; it's much more appropriate for design. The only argument for honours is to make dodgy comparisons with other centres.

Honours degrees were said to be seriously divisive.

I detest honours classification. It's very counter-productive in an area where quality is totally transparent in the folio. We spend four years getting people to work together, developing an understanding of everyone's strengths and weaknesses and then at the eleventh hour we say you scored X and get a 2.1 and you scored Y and get a 2.2. It is very destructive to a vital feature of our School – the positive group ethos.

It is hugely divisive at the end of the year. The people with 2.2s and especially the 3rds are seriously disgruntled. Bosom pals fall out with each other because of the final assessment labelling. It affects their morale. Some go out with their tails between their legs, not feeling positive about their future, feeling they may have missed out by only one mark.

The determination of degree class took too much time. A 2.2 was supposed to be a 'good' degree, but there was no doubt whatsoever that it was perceived as 'having missed out' on a 2.1. Staff felt it was relatively easy to detect the Firsts and also those who probably deserved a Third (which was perceived as having failed, except to endure to the end); it was extremely difficult to draw a dividing line between 2.1 and 2.2 and, after all, in any conventional distribution that was where the majority of candidates would fall: 'We have lengthy arguments about these people every year'. Some staff admitted to having feelings of grave disquiet when these decisions were being made; it seemed unlikely that any assessment system could be sufficiently valid, reliable and precise to make such life-significant decisions with any confidence. After such decisions were made it was of course necessary to display confidence in them – and suppress the misgivings until the following year or until some interviewer encouraged you to talk about them. It was, however, noted by one interviewee that the making of difficult decisions in complex situations should be a defining characteristic of professionalism: being trusted with discretionary judgements was a privilege that higher education teachers earned and from which they should not try to escape.

The fourth objection was not so much to the existence of honours classification, but to its being the only honourable exit point from the course. Only rarely did students get 'counselled out' of the course at the end of year 3; it then seemed quite appalling that weaker students might stick it out to the end of year 4 and then (as if a weakest link) 'leave with nothing'. Such students were given third class honours: 'they ought to have failed and the class of degree they get tells people this'. Some staff were said to believe that the more Firsts that you awarded, the better your course was; thus there had been (it was alleged) a gradual drift in standards. It was now too uncommon for people to fail and too common for them to get better degrees than they deserved. The answer was said to lie on the provision of alternative and honourable exit points (appropriately named certificates, diplomas and unclassified degrees).

It may be wondered if there was such broad agreement about the undesirability of classified honours, why they had ever been instituted in the first place. Four reasons were given: the wish to protect the existence of four year degrees in Scotland; the need to have parity of status with other similar institutions; the 'reactionary intransigence' of the University of Glasgow and the regrettable wish of the more competitive students to have official confirmation of their superiority.

Before moving on from these emerging issues, there are two matters deserving brief comment. Firstly, there was an issue which one person saw as very important: although clearly an assessment problem it was distinctive in that no-one else in this or the other subject areas did more than allude to it in passing. In a situation where there were many more applicants than places available and it was important to end up with the right number of

students, the assessment of applicants' attitudes became very significant – and was always problematic. Information on application forms was of little help; the predictive validity of interviews was limited, especially as some students were becoming very good at convincing the interviewers they possessed desirable attitudes to study and to design. Half and full-day selection workshops had been introduced; these were useful but very staff-intensive.

Secondly there were two matters which had emerged as serious issues in other subject areas but which conspicuously did not emerge in design: the burden of assessment work on staff and plagiarism amongst students. Although it was the case that interviewees commented on the time taken up by assessment, there was no strong complaint about assessment becoming an intolerable burden. Assessment was seen as so important, and indeed so interesting, that it seemed reasonable (on the whole) to make time available for it. Three people commented that refinement of the assessment criteria had actually made the process more focused and thus less time-consuming. A similar number, however, made the point that although time could be found to do assessment, it was very difficult to make time to think about it and to develop it. With regard to plagiarism, there might be some doubt on occasion about the source of ideas, but the nature of portfolio assessment and the familiarity that tutors developed with students' on-going studio work meant that there was very little chance of students dishonestly claiming as their own work which was not.

H. IMPORTED, GENERIC ISSUES

In this penultimate section, five generic issues are addressed. They differ from those identified in the previous section in that the researcher had them in mind before any of the interviews were carried out. They are issues which it was predicted would be of relevance in all the four subject areas to be studied. There were again no direct questions on these matters in the original interview schedules. When people raised these matters, questions of clarification were asked. In what follows the issue is first defined and then a brief summary is given of main points from the data. As this report is already too long, illustrative quotations are kept to a minimum. An apology is offered for the attack of compulsive alliteratitus ('the CAs') in issues two to five.

(a) Assessment: the big purposes.

In the literature, assessment is sometimes described as having three broad purposes: certification, accountability and learning enhancement. At other times the first two of these are conflated and two broad purposes are given: certification/accountability (the control function) and learning enhancement (the growth function).

There was not much direct reference to the accountability purpose i.e. that assessment produced the evidence that courses were doing what was claimed for them and that individual members of staff were doing their job as teachers effectively. However, the degree show was frequently mentioned in relation to public accountability: 'You want to see what we are doing? Here it is.... Judge both students and staff.'

Everyone clearly acknowledged the certification and learning enhancement purposes: both were seen as inescapable. Although there was a little concern amongst a minority of interviewees that some people were getting a higher class of honours than perhaps they deserved, there was no general fear about the quality of student work. Some said it was actually increasing year on year. In other words, the assessment arrangements in place were preserving standards within the School of Design and Craft; the certification function was satisfactorily fulfilled. There was general agreement that assessment procedures had learning

enhancement at their core; the interview evidence supports this claim more convincingly than in any of the other subject areas.

(b) Constructive alignment

Constructive alignment is a term which has appeared only recently in texts and in policy documents on assessment. It refers to that situation in which the aims and objectives of a course (perhaps described as statements of intended learning outcomes), the methods for teaching the course and the ways in which learning is assessed are all planned together and related to each other so that each guides and reinforces the others. Outcomes, methods and assessment then co-exist in a positive, productive relationship i.e. they are constructively aligned.

Only relatively recently have the required approaches to course design emphasised that assessment procedures should be clearly part of the design process. It was unsurprising then that there was more reference to the concept of constructive alignment (if not to the term itself) within the areas of product design and product design engineering (where new degrees have been introduced). The concept is very obviously present in the contribution noted at the bottom of page 21.

What was clearly revealed in all interviews was the existence of what might be called discipline-appropriate alignment. Assessment was planned and carried out in ways which were closely and explicitly related to the perceived nature of the design activity. The predominant underpinning concept was authenticity (in that the assessment of students closely resembled the ways on which qualified design practitioners were themselves judged) rather than validity (in that assessment of students allowed sound inferences about the achievement of explicit course purposes).

(c) Communicating about assessment

Clear assessment criteria are needed for consistency of judgement, but alone they are not enough. What is needed in addition is shared understanding of what the criteria actually mean. Although things like mark schemes, model answers, exemplar portfolios and the like help to attach meaning to criteria, shared understanding commonly comes about by assessors discussing, and indeed arguing, amongst themselves about how criteria are to be applied to actual examples of student work and performance.

The assessment procedures deployed within design were characterised by co-operation and collaboration. This situation should be compared with common current practice in some subject areas within higher education. Individual lecturers set essay topics and examination questions; there is usually some scrutiny of these and revision in the light of comments from informed others, but this appears increasingly to be done by the circulation of drafts by e-mail; in some subjects the feedback that setters receive is minimal. The marking of essays and exam scripts is a solitary business; it is then routinely followed by second and perhaps third marking and any resulting discrepancies are removed by discussion and negotiation. Evidence suggests that pressure on university staff has meant that the resolution of differences has had to become increasingly rapid and economical of staff time. The aggregation of contributory marks has become increasingly formulaic and depersonalised. All of this is very different from standard practice within design; the nature of the process there, as described earlier, means that assessment is something which **must** be talked about at all stages.

Staff, students, colleagues in other specialisms, external examiners, they all talk a lot about the quality of students work over a period. There is a great deal of talking around the products and then around a table – face-to-face over coffee.

The system of assessment develops a dialogue. Five of us assess projects independently, then put them together and debate them. This is a genuine dialogue, not an e-mail interaction.

The geography of GSA buildings means that we talk about student assessment not just around tables, but as we walk along the pavements.

(d) Criterion-referenced assessment

The distinction between norm-referenced assessment (spreading candidates out to make judgements about their relative standing) and criterion-referenced assessment (judging candidates' absolute standing against predetermined criteria) is a familiar one of long-standing. It is important to ask what these things are. They probably represent polar extreme schools of thinking about assessment or even ideologies to which people claim allegiance in an uncritical way. It is doubtful if they can ever be used accurately to describe any actual assessment scheme or strategy. In practice, pure forms of norm-referenced and criterion-referenced assessment do not exist.

Of the four subject areas studied, the two concerned with professional preparation (medicine and design) had moved significantly further towards criterion-referenced thinking and practice than had the others. Within design the requirements of the central registry that percentage marks be submitted and the necessity of awarding classified honours degrees forced some concern for norm-referenced thinking, but for most people (not all) this came quite late in the overall process. (There was one person whose resolution of the norm/criterion referencing tension was very different from the others. S/he seemed happiest in spreading students out over a wide spectrum of marks and then translating these numerical verdicts into descriptive criterion-related statements; this was a most intriguing, perhaps disturbing, reversal of the conventional approach.)

Within the interview data there was also a distinct emphasis on so-called ipsative assessment (where the main concern is assessing one student against that same student's earlier performance). One modification of the assessment profile document had been designed with this specifically in mind and provided good illustrations of students' developing capabilities. There was also a type of assessment thinking that appeared in the contributions of two people that is highly unusual and probably deserves the coinage of a new label—perhaps longitudinal cohort referencing?

We are interested in how a student develops; assessment has both to detect and develop student progression. But in this department we also emphasise year-on-year group progression. Within each year a year-style develops and I want this to be 'better' than the previous year. We make all the previous year portfolios available and say to students that their year has to develop beyond them. This is how we encourage standards to rise from year to year.

(e) Constrained autonomy

The idea of any single member of staff having complete freedom to do what s/he thought best in assessing students is, and always has been, absurd; individual autonomy is always constrained. The idea of a group of academic staff in one subject area having freedom with

respect to how 'their' students are assessed is **not** self-evidently ridiculous. However, even here clear constraints operate and are recognised. Students will be awarded a degree not by one department, faculty or school, but by the institution; thus the institution clearly has to exercise its legitimate authority over what is done in its name. This manifestly includes how results are recorded, aggregated and conveyed to a central registry, but it also includes aspects of quality control. The existence of such constraints on individuals and the design school was seen as entirely proper and reasonable.

Within the School of Design, there seemed to this writer to be acceptance of specialism-based differences in aspects of assessment practice within one broad approach; there was some anxiety, but not a crippling fear, of moves towards uniformity and commonality. It also seemed as if the School of Design was being 'allowed' to be appropriately distinct within GSA from the School of Fine Art although here again there was some anxiety about centrally-derived pressure to uniformity. There were a few references to constraints from the University of Glasgow, particularly as these affected the shared degrees in product design engineering. In product design, there was a recognition that productive partnerships with European institutions required mutual adjustments of assessment practice, but these were not perceived as unwelcome constraints; the benefits outweighed any costs of concession.

As in other subject areas, there was an unease that 'government' and the QAA would probably interfere more in future. There was an expectation that assessment practice would be subjected to more external constraint and direction.

I. CONCLUDING REMARKS

In the opening paragraph three research questions were identified. Why do you assess students in the ways you do? When assessment practice changes, why does it change? What assessment issues emerge as important? Brief answers to these questions, summarising the views of the interviewees, will now be attempted.

- Design students are assessed in the ways described in Section B. Distinctive features of the assessment system are the importance attached to portfolio assessment, the explication of assessment criteria and the showing of student work to an increasing number of assessors (including the public). Important underpinning concepts are authenticity, dialogue and growth.
- In the design area, as in no other subject area explored, staff see assessment as normally occurring **during** teaching/learning and not after it.
- Where new degrees have been introduced significant changes have been made in the assessment of areas outwith design studio practice.
- Within studio practice and critical/historical studies there have been evolutionary changes, mostly concerned with the refinement of assessment criteria and with ways of representing profiles of student performance.
- There has been no radical change in the assessment of studio practice; it must not be concluded from this that there is any need for such change.
- The main factors influencing gradual change have been (a) concern for the quality of student learning and (b) learning from shared good practice in related subject areas.

- There were no significant changes that people wished to make to assessment procedures.
- There were assessment challenges to be faced: the place of the assessment of written work within an increasingly accessible curriculum and the effective assessment of the application to design practice of knowledge and understanding from other areas.
- There was some unease that in future there might be stronger external pressure towards uniformity in assessment practice within both the School of Design and the School of Art as a whole.
- Two issues emerged as salient: firstly, the difficulty of being sure that creativity in design was being validly and fairly assessed and secondly, the lack of justification for the continued existence of classified honours degrees in the design field.
- The abiding impression left with this researcher was of a group of people of goodwill with a strong commitment to developing high quality learning in their students and with a sensitive understanding of how assessment could enhance and inhibit that development. These people were willing to learn from colleagues, but they did not believe that they had adequate time to reflect on their existing assessment practice let alone plan and implement changes.

Colin Holroyd
October, 2001

CHANGES IN STUDENT ASSESSMENT**A Summary Report****Colin Holroyd**

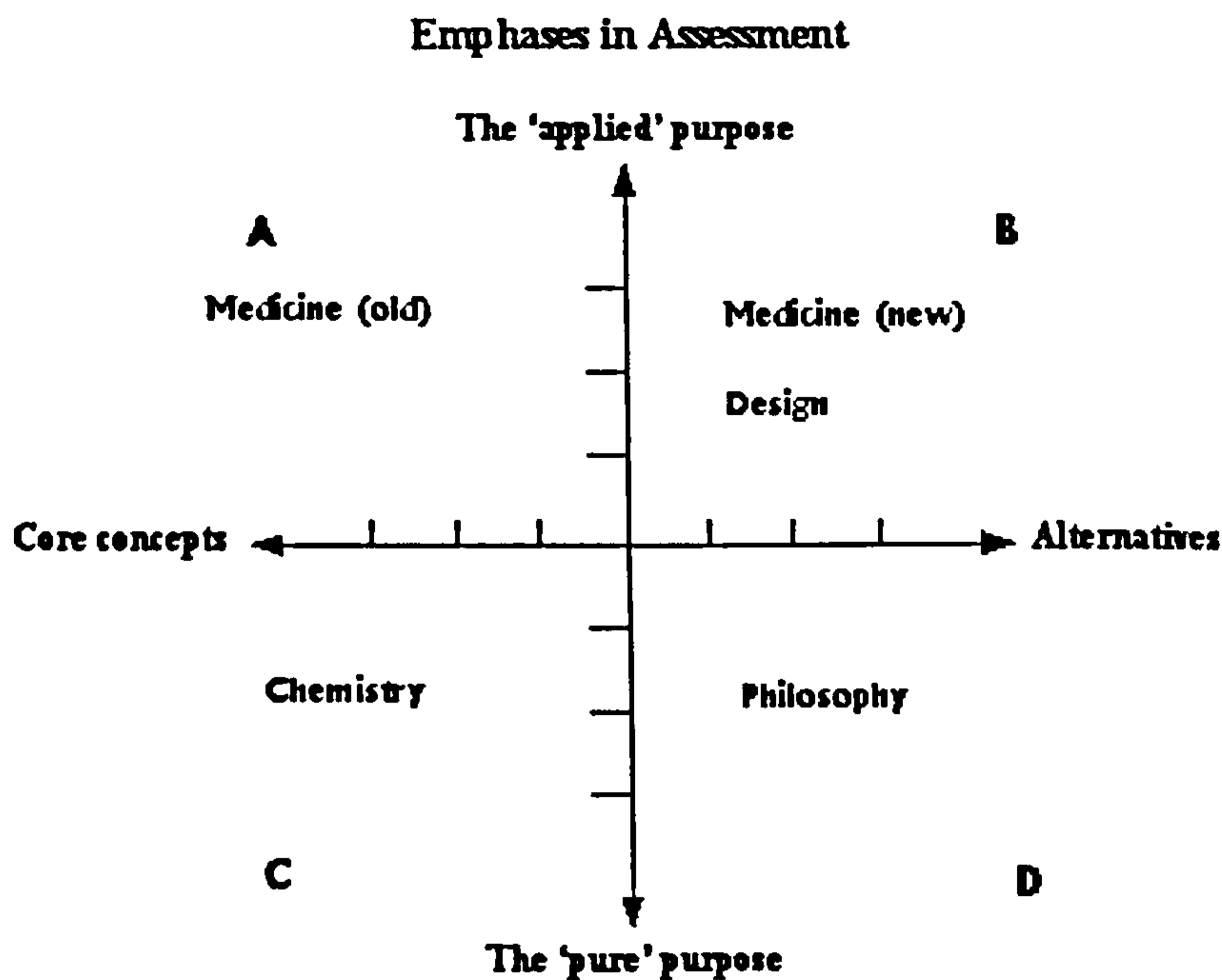
THE PROJECT During session 2000-2001, I carried out interviews with thirty-six members of academic staff from four broad subject areas within Glasgow University (and Glasgow School of Art). This was most of the empirical work for a personal research project exploring the relationships between changing policy and practice in student assessment and changing concepts of academic professionalism. The research aims were all concerned with description and explanation; I had no role as an evaluator of the quality of assessment. In the research reports I do, however, describe how the interviewees evaluate their own practices and in the 'personal comment' and 'general themes' sections of this summary I am probably guilty of straying beyond the data.

A copy of a descriptive report for the relevant subject area has been sent to each person interviewed; a set of the four reports has been lodged with the Vice-Principal (Learning and Teaching) and with the Director of the Teaching and Learning Service (both of whom approved the original research proposal). These reports are quite lengthy and I was asked to produce a short report for circulation. This is that report. Obviously there is much material in the full reports that is not even hinted at here.

I would like to thank all colleagues who gave me generously of their time.

THE SUBJECT AREAS It is not claimed that the four subject areas chosen are in some way representative of the University as a whole; they were, however, chosen to capture some of the diversity within this large, complex institution. It was predicted that subjects would vary along two distinct dimensions related to the purposes of assessing students towards the end of their degree programmes. The first was a pure-applied dimension. Do final assessments emphasise judgements of academic soundness or of fitness-to-practise? The second was a content-process dimension. Do final assessments look for student understanding of defined core content/concepts or for student ability to compare and evaluate alternatives?

Subject areas were chosen in the hope that staff would locate the four subject areas each in a different quadrant of the framework shown overleaf.



The four subjects were chemistry, philosophy, medicine and design. Although staff in all areas were critical of some aspects of this framework, nevertheless they located their subjects as predicted, with the exception of medicine. In medicine, staff were unwilling to locate their subject area in one quadrant only; they said that the 'old' curriculum was in quadrant A; the new (post 1996) curriculum was in quadrant B.

ASSESSMENT IN CHEMISTRY

What was said to be pleasing?

Level 1	The assessment scheme as a whole 'works extremely well'. Short diagnostic tests were 'a very successful innovation'. Laboratory practical work gave good opportunities for informal assessment.
Honours	The final papers were of high quality, with a pleasing emphasis on problem solving. An appropriate weight was given to continuous summative assessment. The research project assessed skills of real value in employment.

Significant changes in the past?

Level 1	Increased attention to continuous summative assessment. Introduction of short diagnostic tests. More emphasis on formative assessment. More serious assessment of practical skills.
Honours	'We haven't tinkered with it much, because it seems to work well.' Single mentions: some increase in weighting to project; the introduction of a carry-over element from Junior Honours assessment; a 'tightening-up' of assessment of lab work.

Changes staff would like to see?

- Level 1** Clearer definition of core knowledge and then more rigorous assessment of it.
More use of objective testing, where appropriate.
More systematic assessment of practical skills.
More effort to achieve common understanding of assessment criteria.
- Honours** Generally 'no – because we already do it well'
Single mentions: assessment of core content; assessment of calculations required by industry; more assessment of oral communication and of structured problem solving.

Changes that may be required of staff?

'Don't really know - and not very interested, but expect there will be.....'

Further intervention from central management; this might be good (for example if it dealt with strategies to tackle plagiarism), but more probably not.

Further 'interference' from outside policy makers; this is fine if from Royal Society of Chemistry, but not if, say, QAA. (Benchmarking 'won't have much effect because benchmarks are written by chemists we know'.)

Emerging concerns?

1. Assessment is a burden; the burden is generally acceptable, but becomes intolerable at peak times.
2. The newer forms of assessment provide more opportunities for student dishonesty than traditional examinations.
3. Multiple-choice questions are disliked, but more use should be made of other forms of objective and computer-based assessment.
4. Examination stress is very real for some students, but it is not clear what the appropriate response should be.

Two personal comments

1. External policy, pressure or influence tends to be unwelcome unless that source shows itself sensitive to disciplinary differences in assessment thinking.
2. More opportunity for communication amongst staff about assessment would increase their appreciation of how assessment impacts on student learning.

ASSESSMENT IN PHILOSOPHY

What was said to be pleasing?

- Level 1** The balance in the assessment by essay and by examination.
The essay allowed early constructive feedback.
The degree exam required students to pull together learning from different sources; it offered less scope for plagiarism.
- Honours** There was a good diversity of methods, appropriately weighted.
The quality of the final examination papers was high.

Significant changes in the past?

- Level 1** No dramatic changes: 'Assessment has always been by essay and examination – and that's not wrong in philosophy'.
- Honours** Summative weighting given to essays.
Introduction of project presentations.

Changes staff would like to see?

- Level 1** Not very much: 'Current situation is pretty good given the constraints which apply'.
However, a few mentioned: more informal and formative assessment; less dependence on essays and exams alone; more originality; reduced reliance on graduate teaching assistants to do the routine work of assessment.
- Honours** Not very much: 'The present situation is sound'.
However, a few mentioned: more informal and diagnostic assessment: more weight to the presentations, more assessment of student ability 'to talk philosophy'; better understanding of criteria for assessment of oral performance.

Changes that may be required of staff?

Expect that extra-departmental influence, pressure, interference will increase; external policy requirements are likely to mean: more assessment of so-called transferable skills;
even more continuous summative assessment (leading to reduced integrity of assessment);
greater economy in assessment to be achieved through use of information technology.

Emerging concerns?

1. The assessment procedures used had progressively reduced originality of thinking by both staff and students.
2. Reduced integrity in assessment inhibits further movement from traditional methods of assessment.

3. An alleged downward drift in honours standard has been reinforced by increased diversity in assessment methods.
4. Others in the department (not those interviewed) complained about the workload of assessment.

Two personal comments

1. The academic philosophy community as a whole could benefit from becoming more creative in its thinking about assessment. The appropriate centre within the Learning and Teaching Support Network could play a useful role in this.
2. The staff in this department do not seem to talk with each other about assessment as much as its importance deserves. Communication is focused on the performance of individual students; time does not seem to be available to discuss the underlying issues.

ASSESSMENT IN MEDICINE

What was said to be pleasing about assessment in the 'new' curriculum?

The new assessment system was coherent and it worked; staff and student reaction was positive.

The MILE (Medical Independent Learning Examination), the OSCE (Objective Structured Clinical Examination) and MEQs (Modified Essay Questions) were effective methods of assessment.

There was a good balance amongst the diverse assessment methods deployed.

There was a significant amount of integrative assessment.

Significant changes from assessment in the 'old' curriculum?

There is now less assessment of factual content.

There is more integrated assessment.

Clinical assessment has become more standardised and thus more reliable.

There is reduced dependence on 'traditional' (and less authentic) assessment methods.

Changes staff would like to see?

Giving students a better idea 'of just exactly where they are'; more useful feedback.

A better mechanism for review of the overall coherence of the assessment system.

Further refinements in assessing clinical capability.

More demanding assessment in Year 4.

Also single mentions: more use of portfolio assessment; computer-based diagnostic assessment; valid assessment of ethical dimensions and of attitudes.

Changes that may be required of staff?

1. Developmental changes from internal monitoring of operation of assessment procedures.
2. Extra-faculty policy and pressure; institution-wide movements towards an undesirable uniformity related to academic rather than professional concerns.

Emerging Issues

1. Assessment of clinical competence and of academic understanding require different types of grading scales.
2. Radical changes in assessment have produced workload shifts which need different management strategies.
3. Not all assessors can be assumed to have adequate assessment expertise.
4. There is a tension between authenticity and reliability in assessment. How can one be increased without decreasing the other?

Two personal points

1. There is remarkable commitment amongst staff to making the new assessments work; for this to be retained staff need more time to think, to communicate with each other and to implement further agreed modifications.
2. The new assessment system emphasises the need for a mechanism for monitoring the operation of procedures, for initiating review and for encouraging the ongoing professional development of those who assess.

ASSESSMENT IN DESIGN (Glasgow School of Art)

What was said to be currently pleasing?

The assessment scheme required and promoted regular dialogue with students.

The emphasis on portfolio assessment for studio practice gave the assessment scheme high authenticity.

The generally applicable assessment framework (specifying criteria and levels) was valuable and had economised on staff time.

Any potential for unfairness to students in those aspects of design where judgement could be highly subjective was minimised by staff collaboration.

The assessment process culminated in a highly public event.

Significant changes in the past?

‘Things have not changed fundamentally – and I don’t think they should have done.’

However, a few mentions of – changes in the wording of assessment criteria, their relative weightings, the meanings attached to certain words and the ways in which assessment profiles are portrayed.

Changes staff would like to see?

‘No change that is really urgent....’

However, a general request for more time (to explain assessment to students, to provide feedback, to think about changes...). Plus a few mentions of – greater comparability across sub-specialisms, reversal of some downward drift in standards, more ‘non-standard’ approaches in the assessment of work in academic areas (Historical and Critical Studies), better assessment of the **application** to design tasks of contextual and theoretical understanding.

Changes that may be required of staff?

‘I don’t anticipate or fear any major pressure or upheaval.’

However, a few mentions that central management are becoming over-keen on uniformity and that more emphasis on numerical and quantitative approaches within assessment is leading to more formulaic decision-making in aggregation and degree classification.

Emerging Issues?

1. There is one huge, dominating issue: what is the place of creativity in design and how is it to be validly and reliably assessed?
2. There is little justification for the existence of classified honours degrees within design.

Two personal points

1. In this subject area, as no other studied, assessment is perceived as integral to teaching; it is informed by regular observation of students at work and by regular talk with students. This allows staff to have great faith in what they call ‘descriptive categorical assessment’ and considerable disdain for marks and numbers. The only obvious down-side to this is a somewhat ingenuous belief that consensus in assessment is the same as validity.
2. Assessment was not seen as a burden; however, staff did not believe they had adequate time to reflect on their assessment practice or develop it.

GENERAL THEMES

Each of these summary accounts has indicated the issues which emerged as important from what the interviewees had to say. It may be helpful now to indicate four broad themes which come from some reflection on the data from all the four areas. This is a personal and provisional selection; more work will be done in this current session.

The extent of change – a cautionary note.

The nature of changes – an amplification.

The encouragement of change – some implications.

Assessment and academic professionalism

(a) The extent of change – a cautionary note.

It is now common to read that in the last ten years or so higher education has been the site of unprecedented change. For example, Becher and Trowler (2001) have recently outlined the far-reaching, inter-related sets of changes in the character and activities of higher education due to post-industrialism, globalisation, massification, altered HE-state-industry interactions, managerialism within universities and substantive disciplinary growth. Within this research there was an explicit focus on change in assessment practices in one institution of higher education. It would be tempting, but too easy, to leap to a judgement that assessment practice in a subject area must currently be unsatisfactory unless it has recently undergone some radical change. Evaluation of the current quality of assessment practice should not merely be based on the existence of recent changes.

This summer a leaflet landed on my desk advertising a staff development course in assessment for higher education staff. It contained this assertion: 'It is obvious that the traditional methods of assessment which were adequate in the past are no longer good enough and that things have to change'. This present research did not proceed on the assumption that such a conclusion was obvious.

In three of the four areas studied, there has been a modest amount of change in assessment practice. The research was not designed to produce evidence for a judgement that current practice was 'not good enough' in those areas, or indeed that assessment in the fourth areas was somehow 'better' than the others because of the very significant changes recently implemented. It is of course possible (if highly unlikely in the case in point) for change to be unhelpful and make things worse.

All of the following points were made by those interviewed. 'We have introduced many changes, but in areas other than assessment.' 'Our assessment system works well; there has been no compelling reason for fundamental change.' 'There have been widespread changes in assessment generally; that doesn't mean change was needed in every specific location.' 'Gradual evolutionary change is much better than any radical shift.' 'We were way ahead of the game twenty years ago; others may have caught up a bit – but we are still OK.' 'The context of assessment has changed; because the conditions are now much more difficult, the criteria by which adequacy of assessment is judged have been relaxed a bit.' The research does not allow any comment on the credibility of these claims.

(b) The nature of the changes – an amplification

In the assessment literature, changes in assessment practice tend to be seen as being of two types. The first are discrete innovations within existing courses introduced through the efforts of one or two enthusiastic individuals. Many of the assessment changes described in the

ASSHE Inventory (1996) are of this type. The second are broad evolutionary trends involving gradual shifts in the dominance of different assessment methods. Several texts, for example Freeman and Lewis (1998), give lists of examples of such trends.

The changes described by those interviewed in this research suggest that a four-fold classification of assessment changes may be more useful, especially as the four types seem to be associated with different patterns of causal factors. Firstly, there were clear examples of locally-based innovations (the first type above); the introduction of the four short diagnostic tests in chemistry was just one example. This came about because of the conviction of a few staff that student learning could be enhanced by a different assessment method. Secondly, there were broad evolutionary trends; these were attributed to the influence of educational fashion and institutional climate. ('They seemed to be favoured by the people up the hill'). The increased weighting attached to continuous summative assessment was the obvious example here. The third type overlapped a little with the second, but examples had a distinctive character in that they were attributed to external, policy-directed shifts. The source of pressure was perceived as extra-institutional (vaguely 'the government' and more precisely 'bodies like QAA'), but channelled through central management. The priority to be attached to key, transferable skills was an example of this; it had required more attention to the assessment of things like oral communication and presentation skills.

The most dramatic changes in assessment practice, however, were not of any of these three types. They can perhaps be labelled as new-course related discontinuities. In three of the four areas explored, radical change in assessment practice had occurred because new degrees (or completely new degree programmes) had been planned and introduced. As part of the course design process that is now required, it had been necessary to re-think assessment systems from first principles. This had resulted in fundamental changes in the assessment of chemistry (in the MSci), in design (in BEng, MEng, BDes and MEdes) and in medicine (in the 1996 curriculum). Changes of this type all involved a diversification in assessment methods; this stemmed from two official demands (a) that the intended learning outcomes be made explicit and (b) that assessment methods be constructively aligned with such stated course purposes. The acknowledgement of diverse purposes and an acceptance of the demands of validity in assessment leads inevitably to diversity in assessment.

If there is a message here, it is that if academic leaders want changes in assessment practice, they are more likely to get them through the ab-initio planning of new courses than from modifications of existing ones.

(c) The encouragement of change – some implications

Accepting entirely that change for change's sake is not a good thing, one can still argue that improvement in assessment is possible and, in some cases, highly desirable. The perfect assessment system has not yet been devised (and never will be); in no subject area within this university is assessment so good that it cannot be bettered. How then might the improvement of assessment be encouraged? There are some useful pointers from the interview data.

(i) The evidence suggests that all lecturers are highly conscientious in their assessment of student work and that a few (quite often the younger and less experienced ones) have strong views and creative ideas about assessment change. Unfortunately it also seems that proposals for change are not always welcome. ('That's not the way we do things around here; if it isn't broken, don't fix it.') There is a role here for heads of departments and for the chairs of departmental learning and teaching committees to ensure that the voices of younger staff are heard and that their views are carefully considered.

This research was quite intentionally focused on staff; it did not seek any student perspective on assessment. The point just made about staff should be extended to cover students as well.

(ii) In one area in particular it was very clear that positive developments in assessment had come about from the members of one sub-specialism listening to, and learning from, staff-in related sub-specialisms. This suggests an important role for whole-faculty learning and teaching committees. For such groups it should be made wholly explicit that 'learning and teaching' is a short-hand term for 'learning, teaching and assessment'.

(iii) It was very noticeable that staff were suspicious of, and resistant to, advice and pressure coming from any source seen as insufficiently aware of disciplinary differences in assessment thinking. Two examples must suffice. Chemists take very seriously any guidance on assessment coming from the Royal Society of Chemistry, but tend to reconstruct (at best) or circumvent (at worst) any pressure from 'The University' or the QAA because 'they do not understand the nature of our subject'. Medical staff accept advice, indeed direction, from the General Medical Council because 'after all they are themselves medics'.

This means that any university-wide policy generating group has to proceed with considerable care. Interviewees frequently said that to-date there had been relatively little external policy, or central steering that had had great impact on assessment practice; they also said that they anticipated that in the future policy requirements would become more extensive, invasive and prescriptive.

It was accepted that there should be university regulations on assessment that clearly defined, for some aspects of assessment, what must be done and what must not. This amount of prescription and proscription was seen as entirely legitimate. There also seemed to be a recognition that insofar as the university had some unity of purpose, which inevitably involved a focus on academic standards, then there should be agreement on those normative principles which should underpin all good assessment practice within the university. One of these principles should be that there are crucial aspects of assessment that can only properly be decided at faculty and departmental level. The difficulty of course lies in determining in what areas of assessment there should be university-wide uniformity and in what areas there should be subject-related diversity.

The interview data supported experience elsewhere that centre/periphery and top-down models for achieving assessment development have limited chances of success.

(iv) Assessment and academic professionalism

I have a strongly held opinion that insufficient thought has been given to assessment as a professional activity of academics. I have made a modest contribution to thinking on the topic in a recent paper. (Holroyd, 2000)

I was much struck by a remark made by a head of department when I was negotiating access to staff: 'We will happily co-operate in your research, but it does depend when. At some times in the session we are far too busy doing assessment to think about it'. In all the subject areas interviewees regularly said that they could find time to carry out the work of assessment, but they could not find time to talk about it and to think about it, let alone devise and implement changes to the existing scheme. The sheer burden of assessment inhibited innovation. If we adopt the currently fashionable (and facile) definition of the professional as reflective practitioner, then some doubt arises whether practice with no time for reflection can be genuinely professional. It could be argued that the actual process of assessment requires 'reflection-in-action' and that the ability to do this is an aspect of professionalism. Even so, is there time for reflection-on-action?

That simple conceptualisation of professionalism is inadequate. What if we consider the standard defining characteristics of traditional professionalism? Firstly, professionalism requires a well-developed knowledge-base. Most of the interviewees had extensive knowledge of assessment derived from personal experience (both of assessing and being assessed), 'common-sense' and folk-wisdom; a few had read widely and deeply in the subject. It was notable how many people said that they did not know enough about what might be called the scholarship of assessment and that they wished to learn more. It was also striking that one person declined to take part in the research because 'although I have responsibility for the assessment of X, I have no specialised knowledge of assessment'. Secondly, professionalism requires adherence to a clear ethical code. Ethical requirements are probably being met by the majority of individuals and departments, but it has been recognised by the university that the existing regulations for assessment are incomplete, at times inconsistent and out-of-date and the production of a new Code is in hand. Such a Code will be essentially a legalistic document but it ought also to contain clear ethical specifications. Thirdly, professionalism allows people to be entrusted with the making of discretionary judgements within complex areas. The assessment of student abilities and achievements and deciding the appropriate levels of award is undoubtedly a complex business and the importance of informed professional, academic judgement is generally recognised. It is, however, considered by some to be under threat due to pressing economies of time and from increasingly impersonal, formulaic and mechanistic processes in aggregation and in degree-classification decisions.

Those writers who deal with re-defined and emergent professionalism again stress explicit ethical requirements, but they add to this new commitments to enhanced communication (with fellow professionals), to negotiation (with clients) and to partnership (with other professional groups). The place of negotiation within assessment is problematic, but well deserves clarification. The problems of partnership, but also its potential contribution to professional growth, are increasingly prominent in the more vocationally-oriented areas of medicine and design. The requirement that there be regular communication with fellow assessors is the most worrying; interviewees commented frequently on the shortage of time and of appropriate locations for formal and informal discussion with colleagues about issues in assessment.

Overall the need for professionalism in assessment, however defined, was clearly recognised. The level of professionalism which currently exists was thought to be under some threat; the conditions for increasing professionalism were thought not to be favourable.

It is perhaps too easy for me to suggest that the main obstacle to the maintenance and enhancement of assessment professionalism is simply a shortage of time in the lives of academics. It is perhaps the case that academic teachers do not wish to reflect more on their assessment practice. There may be risks in being seen to devote more attention to reflection on the educational aspects of their work; some academics wish to continue to 'borrow' their professional status from the professions which they prepare their students to enter rather than seek to acquire the dual professionalism open to academic teachers, as both academics and as educators. The demands of reflective professionalism as an assessor/educator may be seen as making mastery of the 'home' discipline less likely. Although Bourdieu was writing of higher education in France in 1968, his remarks are still unsettling.

..... owing to the particularly archaic nature of an educational system hypnotised by illusory images of its grandeur, those consecrated by a bankrupt institution were obliged..... to invent new ways of playing the part of the teacher (all based on adopting a reflexive distance from practice), by lending him [sic] the strange features of an intellectual master of reflection who reflects on himself and in so doing helps to destroy himself qua master. (Bourdieu, 1988)

SUMMARY POINTS

- (a) The quality of current assessment practice should not be judged by the extent of recent change.
- (b) Radical change in assessment comes through overall course planning.
- (c) Improvements in assessment practice need more time for staff to talk and to think about assessment.
- (d) Pressure to change is more welcome when it shows awareness of subject-based differences in assessment thinking.
- (e) Professionalism in assessment is under some threat.

CONCLUDING NOTE

It will be obvious from the above that there are significant and varied opportunities for the Teaching and Learning Service (a) to support individuals, departments and faculties in maintaining and enhancing the quality of assessment practice and (b) that short courses, award-bearing courses, out-reach departmental consultancy and membership of working groups and committees all have a useful part to play.

REFERENCES

- ASSHE (1996) *The ASSHE Inventory: Changing Assessment Practices in Scottish Higher Education* Hounsell, D. McCulloch, M. and Scott, M. (eds)
University of Edinburgh: Centre for Teaching, Learning and Assessment
- Bourdieu, P. (1988) *Homo Academicus* Cambridge: Polity Press
- Becher, T. and Trowler, P.R. (2001) *Academic Tribes and Territories: Intellectual Enquiry and the Cultures of Disciplines* Buckingham: The Society for Research into Higher Education and Open University Press
- Freeman, R. and Lewis, R. (1998) *Planning and Implementing Assessment* London: Kogan Page
- Holroyd, C. (2000) 'Are assessors professional? Student assessment and the professionalism of academics.' *Active Learning in Higher Education* Vol 1, No 1, pp 28-45

Colin Holroyd
Teaching and Learning Service
University of Glasgow
October 2001

UNIVERSITY OF STIRLING

Edd PROGRAMME

UNIT 3 ASSIGNMENT

**THE IMPACT OF EMERGING STUDENT ASSESSMENT POLICY ON THE
PROFESSIONALISM OF ACADEMICS**

Colin Holroyd

January 2000

THE IMPACT OF EMERGING STUDENT ASSESSMENT POLICY ON THE PROFESSIONALISM OF ACADEMICS

1. FOCUS ON POLICY

1.1 A policy on student assessment?

What is the policy on student assessment? It may be that there is never a simple answer to questions of this type; it is certainly the case with student assessment in higher education that it is particularly difficult to get near any satisfactory answer. There are perhaps three main reasons for this. Firstly, assessment policy is currently emerging from a number of national sources; there is not one clear, certain policy defining the policy context for institutional operations and professional practices, but rather a firmament of potential policy influences. Secondly, the emerging policy from 'ecclesia' (i.e. the public sites of policy determination) is frequently that policy must be formulated in the agora (i.e. the meeting places where ideas are traded - the separate institutions of higher education) and must be such that it is in line with the central guidelines. Thirdly, not only are the eventual forms of the policy as yet unclear but also these forms will themselves be temporary. A relative impermanence is implied even in some vertical, rational-technical model of policy-making; it enters when feedback from the evaluation of policy implementation influences policy-reformulation. However, the impermanence is more obvious when we think of the horizontal dimension of policy - the interactions amongst policy participants in different organisations outside some line of hierarchical authority. Those living in the more private sites (oikes) of our higher education institutions will not merely receive policy, they will redefine, reconstruct or resist it. Not only does the policy influence the practice, but the practice influences the policy.

These descriptions of policy as vertical and horizontal are derived from Colebatch (1998 - pp 37 - 39); the same author would contend that the question with which this essay began is quite simply the wrong question to ask. For our present concerns a better question might be: What appear to be significant foci of policy activity within the beleaguered space of student assessment in higher education?

1.2 Policy emerging from central sources

Student assessment is changing, and will continue to change, because of the recent and continuing emergence of national policy on assessment. Although it would not be true to suggest that such policy has only been detectable in the last two or three years, it is only since the report of the National Committee of Inquiry into Higher Education (NCIHE, 1997) - the Dearing Report - that this has been an obvious feature. The Committee made recommendations to Government and, of course, recommendations do not automatically become official national policy, as the issue of student fees makes very clear. Nevertheless it can be predicted with confidence that the report as a whole will have a significant influence on the general thrust of policy activity including that on assessment. Developments are ongoing, so what follows is necessarily tentative.

¹ The terms ecclesia, agora and oikes were used by Jon Nixon in a lecture on 3.12.99 and come from Baumann's 'In Search of Politics'.

There is a seeming paradox. The Dearing Report attaches a high priority to the promotion of learning and the enhancing of standards of student achievement. Why then in its 88 main recommendations is there no explicit mention of student assessment at all? (The word 'assessment' is only used in relation to the Research Assessment Exercise in recommendations 32, 33 and 34.) The full report (NCIHE, 1997 Vol 1) does make explicit reference to student assessment and directly addresses the topic - on pages 137-139 - but oddly this section results in no recommendations. Does this demolish any claim that the Dearing Report will significantly affect assessment practice? The paradox disappears if we decide that Dearing saw assessment as a part of any effective strategy for the management of learning - perhaps one that seemed so obvious that explicit reference was judged unnecessary.

If this view is taken of assessment, then several recommendations have significant implications for student assessment. Noteworthy are: the priority to be attached to promoting student learning (recommendation 8); the training and accreditation of university teachers (recommendations 13 and 48); the development of programme specifications (recommendation 22); the amendment of the remit of the Quality Assurance Agency to include standards verification and the development of a code of practice (including a section on external examining and student assessment) to be adopted by all institutions as a condition of continued public funding (recommendation 24) and the setting-up of small expert teams to provide benchmark standards/statements (recommendation 25). Several of these topics deserve more extended discussion and comment than is possible here. [See, for example, Smith et al (1999)] Some quick comments, however, are required on benchmarking and on the code of practice in relation to external examiners and student assessment.

Benchmark statements will define for the United Kingdom as a whole, and for the first time, what has to be assessed within honours degrees in forty-two different subject areas. It is important to ask just how far the prescriptiveness of these benchmark statements will extend. Firstly, the statements apply to honours degrees within a discipline, but they will have a major impact on all those courses through which students progress to honours level and then indirectly on the combinations of these courses into multi-disciplinary degrees. Secondly, it is made wholly explicit that statements are not to define subject content - this is to remain a matter for individual institutions. There does seem to be just a whiff of disingenuousness here; it is not obvious that understanding within a discipline (which is to be included in benchmark statements) can exist independently of the subject content which is to be understood (which is not included in benchmarking). Thirdly, the statements do prescribe what is to be assessed, but not how it is to be assessed; however, such a prescription taken along with the exhortation that assessment procedures be valid does exert a strong influence on methods of assessment.

Recommendation 24 of Dearing concerns the development of a suite of inter-related documents which together will form a code of practice which will be adopted by all universities. The section of this code devoted to external examining has already been published. (QAA, 1999) Rather than look at the substance of the document (which is unsurprising) it is instructive for present purposes to consider its form. The document articulates a series of nineteen precepts each of which has accompanying outline guidance. The precepts 'identify those key matters which the Agency expects an institution to be able to demonstrate it is addressing effectively through its own quality assurance mechanisms'. The guidance is meant to be neither prescriptive nor exhaustive, but to offer a framework which individual institutions may wish to adopt or adapt to meet their own cultures; 'nevertheless, in many institutions the guidance will constitute appropriate good practice'. If a

code is a set of rules or a body of laws, if a precept is a command, rule or instruction and if guidance defines what should 'normally' be seen as good practice, then it seems we do have here a national policy which it will be difficult for any university to ignore.

The document dealing with student assessment in general is expected to be published in the second half of 2000. Informed sources suggest that if publication is to be as early as this (as the Government is said to wish), then the substance is likely to be very similar to that of the Guidelines on Student Assessment produced by the QAA's predecessor, the Higher Education Quality Council. (HEQC, 1996) It seems likely that the form will be the same as that of the section on external examining i.e. a set of precepts with accompanying guidance as to what constitutes good practice.

1.3 Main foci of policy activity

The central concerns of policy development may thus be predicted to be as follows.

- (i) Programme specifications will have to show that intended learning outcomes and the assessment procedures to show the level of student achievement of these are positively aligned.
- (ii) Programme and course specifications and assessment procedures will have to be modified to conform to the implications of nationally formulated benchmark statements.
- (iii) Nationwide expectations about the comparability of standards will require to be met; institutions will be required to demonstrate that their locally determined assessment procedures effectively address nationally defined precepts.
- (iv) Both internal and external examiners will require to receive such training, support and continuing professional development as is needed for their activities to meet the necessary quality standards.

2. FOCUS ON IMPACT OF POLICY

2.1 Two metaphors for impact

The word impact suggests a physical metaphor, one based on snooker. A coloured ball is at rest on the table; a moving white ball hits it and the coloured ball starts moving. Of course two moving balls could impact on each other, in which case the direction and rate of movement of both balls will be changed by the impact. The observable changes are in velocity; less obviously the snooker balls are changed in nature; there has been some deformation of both; old snooker balls have eventually to be replaced. The less perfect the collision, i.e. the longer the two are in contact, the greater the deformation. The metaphor can be extended: the coloured ball may move into a pocket - an outcome of some value; it may bounce right off the table; the coloured ball can snooker others. Everything is complicated further when spin is introduced.

If a physical metaphor of snooker balls inadequately represents people, so too does a rational-scientific model inadequately capture the nature of policy impact. Shift then to a second metaphor

where the focus is on the players rather than the instruments of play and note in passing the retention of some post-modern sense of ludic lightness. In this metaphor, the play is the thing: a dramatic event in rehearsal and performance. There are many different versions of this basic metaphor. In one, there is an authoritative script and the director begins with a clear vision of what would constitute an accurate staging of the drama; in another there is an idea but no script to begin with as the individual actors seek to create an event which is authentic for them and then the director produces a script to reduce the chaos s/he perceives. In both cases the actors come from the relatively private spaces of home and dressing room to form a company in a more public arena. Each interprets the script; lines are forgotten, modified and substituted; occasionally an actor takes fright and cannot go on or, once on, freezes; there are frequent status and ego conflicts. The script is reconstructed rather than implemented; the actors are never merely players. Nor are they characters without a personal history; they have been trained and they belong to Equity, an organisation which influences them from a distance with definitions of what is considered acceptable performance. No two performances are the same; they alter in response to audience reaction and even to the views of the backers or stakeholders. The official script influences the performance, but in the performance there are only personal, internal scripts. The drama is a complex agonistic event and the stage is a contested terrain. And of course most of the intended high drama has elements of low farce.

There are no academic references for the above. The present writer believes the metaphors to be original, which may only mean that the sources are long-forgotten. It is not original to depict vertical policy-making as a tidy rational process (Colebatch, 1998 - pp 106-107) nor horizontal policy-activity as a game. (Was it Damon Runyan who described it as a floating crap-game?) The second metaphor has echoes of Ozga (2000) and even of Foucault on policy as....

...a contact point where techniques of domination - or power - and techniques of the self interact, where technologies of domination of individuals over one another have recourse to processes by which the individual acts upon himself and conversely,where techniques of the self are integrated into structures of coercion. (Foucault, 1980, quoted in Burchell, 1993 - p268)

2.2 The impact of policy on people

One, albeit limited, view of the impact of policy (national or organisational) is obtained from an evaluation of the success of its implementation. 'Policy X required certain actions. Were these actions taken? What other actions were taken?' 'Policy Y required this pattern of resource allocation. How were resources actually allocated?' 'Policy Z was intended to achieve certain outcomes. What were the actual outcomes?'

This can usefully be supplemented by studying how the policy affects what all the policy participants do (and by policy participants is meant the leaders, the aides, the followers, the maintenance staff, the interested parties and the knowledge workers) and how they feel about what they are required to do. From a substantial literature just one example will be quoted, one pertinent to the subject of this essay. How does current policy-change impact on academics? From his research, Trowler (1998 - chapter 5)² concluded that academics fall into four groups depending on whether (i) they like the central thrust of the policy change (the content-discontent axis) and (ii)

² Features of Trowler's analysis have been criticised by this writer in an extended review of the book referred to. This was published in *Teaching in Higher Education* Vol 3 1998

they view policy as something to be implemented or something to be changed (or worked around). If the policy is liked and accepted, then people swim; if it is disliked and accepted, they sink; if it is liked, but regarded as changeable, then policy is actively reconstructed; if disliked and regarded as mutable, the impact results in different coping strategies - of policy avoidance or policy-demand reduction.

3 FOCUS ON ASSESSMENT

3.1 General patterns of change

The coloured ball of assessment practice was already moving before the white ball of national policy impacted upon it. (The temptation to stray into a dental metaphor about impacted wisdom will be avoided.) Or, rather better, the institutionally based actors have been developing their local scripts before the official version is published.

It is normal for texts on teaching, learning and assessment in higher education to give summary accounts of how assessment practices have changed over the last couple of decades. These accounts are not of course identical, but there is substantial overlap. To take just four examples, important features of change are identified by Biggs (1999, chapter 8), Brown and Glasner (1999, chapter 2), Freeman and Lewis (1998, page 310) and Toohey (1999, chapter 9).

Important dimensions of change are said to include the following:

- (i) increasing emphasis on the learning enhancement purpose of assessment rather than its certification and accountability purposes;
- (ii) increased attention to formative rather than summative aspects;
- (iii) more emphasis on a standards model of assessment, involving criterion referenced assessment, and less on a measurement model, involving norm-referenced assessment;
- (iv) more frequent provision of descriptive comment and constructive feedback and less restriction of assessor response to marks, grades and summary labels;
- (v) a move from dependence on one main method of assessment (and end-of-course assessment) to deploying a variety of methods (and within-course assessment);
- (vi) less reliance on assessment by teaching staff alone and more involvement of self, peers and workplace assessors;
- (vii) increased insistence on assessment as integral to teaching rather than a separate activity occurring after teaching.

Of course, any such list immediately prompts questions and debate; these will not be pursued here. It should just be noted, however, that these dimensions of change are not separate and distinct; several of them are closely related and (iii) is at a higher level of generality and includes some of the others. Also, to identify these as system-wide trends has to allow that there will be local contravention of such trends; there are no doubt many places and courses where change has been in an opposite direction. There are also some pervasive countervailing pressures which could result in general reversals. For example, it is frequently claimed that validity is in the ascendant in assessment thinking and activity; this is evidenced by the move to more authentic forms of assessment and the

insistence that assessment procedures be more constructively aligned with statements of intended learning outcomes. On the other hand, others would argue that pressures for accountability, transparency and consistency from government and from potentially litigious students-as-consumers mean that reliability is a more compelling requirement than ever before.

3.2 Change related to earlier policy on modularisation

There is a tendency within the literature for it to be assumed that the important decisions about assessment are all to do with selecting the method and then marking the products. This leaves out a great deal; crucially it omits the management of assessment and decisions on aggregation and final awards. Both, particularly the latter, have been profoundly altered by relatively recent moves towards the modularisation of higher education provision. The impact of these is helpfully analysed by Wolf and her co-authors in the report 'Assessment in higher education and the role of "graduateness"'. (HEQC, 1997)

Wolf distinguishes two ways of modelling how universities assess students' work and make decisions about degree awards. In the first, the bulk of assessment is by end-of-year examination. There may be some options, but the course is not part of a full-scale modular system. The examiners (both internal and external) know each other well and most have considerable expertise in examining this kind of degree. Exam scripts are usually double-marked blind. There will probably have been a meeting of the examining board to scrutinise questions and marking criteria. All examiners attend the board and final decisions on awards and classes are made there. There is the opportunity for shared understandings about assessment criteria and standards to develop. This 'type' has not yet disappeared at undergraduate level, but it applies to a decreasing proportion of undergraduates. Many sub-sections of degree programmes approximate to it and it is still typical at postgraduate level. Twenty years ago it would have been far more common and it is the approach into which most academics were themselves socialised.

The second model typifies the awarding process for modular degrees in which many subjects are taken as options and in which many 'home programme' students have selected options from a range of other programmes. Not all work has been double-marked. Written criteria may not be drawn up by the programme teams. There have been no scrutiny meetings. For any one given student, many of the staff making assessment decisions will have had little experience of examining together. Only a small number of examiners attend the examination boards; final decisions may well be taken later at programme and faculty boards. This second type encapsulates trends in higher education which are recent, but which currently show no signs of reversing themselves.

The emergence of this second model is hugely significant. All the available research evidence points to the great importance of assessors being members of a communicating network; only in this way do similar meanings get attached to the standards implicit in assessment criteria. When few examiners are involved in programme or degree-wide decision-making fora, then their socialisation has been into other groups. They (and external examiners) will necessarily be presented with assessments from subject areas in which their own expertise is weak and with issues of aggregation which are technically insoluble. Under these circumstances their predominant concerns will have to be for procedural correctness and 'due process' and their decisions will be largely formula-driven. These features clearly relate to the professionalism of assessors.

3.3 The likely impact on assessors of emerging policy

There may be a time in the future when it is possible for research to answer: ‘What was the impact of these assessment policies on the work of university teachers/assessors?’ Attempting to answer questions about the likely impact of emerging policy has to be a matter of extrapolation and prediction.

It is common to ask what assessment procedures require of those who are assessed; it is rare to ask what assessment requires of those who do the assessing. This writer has recently attempted to suggest answers to this latter question; the³ result is much too long to report here in detail so a “ quick outline of the process and nine main conclusions will have to suffice. Student assessment is not one activity, or even a set of activities all similar in type. It is a complex business which can involve, it is suggested, eleven types of activity from the design and management of assessment systems, through the creation of assessment situations and instruments and the marking of student responses to them, to the making of decisions about awards in boards of examiners. General trends in assessment (3.1), past (3.2) and emerging (3.3) policies (both national and institutional) all have differential impacts on each of these types of activity; each influence alters what it is desirable for assessors to know and be able to do across the range of aspects.

The main conclusions are as follows.

- (i) Most aspects of assessment will continue to demand discipline-specific content knowledge. The importance of this must not be under-estimated.**
- (ii) For a few aspects of assessment what can be called ‘assessment craft knowledge’ will be adequate for assessors ‘to get by’.**
- (iii) Some aspects of assessment will increasingly require generic ‘assessment scholarship’.**
- (iv) In many aspects both institutional and personal autonomy, as traditionally defined, will be significantly reduced.**
- (v) Effective assessment will increasingly require to draw on an extended range of different knowledge bases; single individuals will be unlikely to possess all the understandings required, so the necessity for co-operation and team-working will become more obvious.**
- (vi) Consistent utilisation of benchmarking statements will require assessors within a discipline to communicate more effectively with each other; this will reinforce the existing message that reliability in assessment comes from shared understanding and is only achieved when assessors talk regularly with each other.**
- (vii) In most, but not all, aspects of assessment decisions will remain difficult, complex and discretionary; sensitivity to the moral dimensions in these decisions has to remain an important feature.**

³ The Director of the Teaching and Learning Service, University of Glasgow, became aware of this endeavour and thought it potentially useful for planning the future provision of support to academics. The document is still evolving and should be available later this year.

(viii) Assessment must be seen as an arena requiring **critical activity**, one in which the purposes, not only of courses and programmes, but of higher education itself and the role of participants within it are the objects of critical thought and action. Realising quality in assessment means an awareness of its super-complexity. (Barnett, 1997 and 2000)

(ix) The need for all involved in assessment (whether graduate teaching assistants, new or experienced lecturers, course managers and central administrators or external reviewers and examiners) to be adequately trained and supported through continuing professional development will be even more obvious.

4. FOCUS ON PROFESSIONALISM

4.1 Changing characteristics of professionalism

Some analysis of professionalism is inescapable; the problem is that professionalism is not some social-scientific absolute, but a historically changing and socially constructed concept-in-use. The extensive literature variously describes professionalism as traditional, quintessential or residual; as semi-, restricted or extended; as dispersed, hybrid or syncretistic; as corporatist or bureaucratic; as reimagined, redefined, retheorised, reconceptualised, regenerated, reborn or reclaimed.

There is some measure of agreement that traditional public-sector professionalism had four defining features.

- (i) Professionalism was made manifest in the pursuit of an occupation or calling which provided the person with a living (as distinct from amateurism).
- (ii) The members of a profession had completed both some form of higher education which involved the mastery of some difficult body of useful knowledge and some extended form of training in relevant skills.
- (iii) The primary allegiance was to a self-controlling professional body rather than to an employer; this body acted as guardians of the specialist knowledge, gatekeepers to membership and enforcers of an ethical code.
- (iv) The substantial gap between the knowledge and competence of the professional and of the client and manager meant it was difficult for these others to evaluate the competence of the professional and thus professionalism necessarily implied a degree of autonomy, of self-control and of trust.

Freidson (1994) effectively summarises all this into two basic requirements of professionalism: firstly the taking of complex, discretionary decisions should be followed by effective action (i.e. the requirement to do good work) and secondly that the work should have a clear value for the well-being of society as a whole (i.e. the requirement to do Good Work).

Traditional professionalism entered a period of crisis in the 1980s. A number of reasons are offered for this: the debasement of some professions into self-serving cabals; the proliferation of groups making flimsy claims to status for financial gain; the fragmentation of existing professional groups; growing public distrust of experts; growing managerialism, accountability and state control and a re-theorising of the status of specialised knowledge. The period since has witnessed the evolution of new forms of professionalism which may be better adapted for survival in a post-modern

environment. For example, Nixon et al (1997 a) would see a regenerated professionalism as necessarily involving collegiality (amongst the members of the profession), negotiation (between professionals and clients), cooperation (with members of other professions) and partnership (between professionals and other stakeholders). This new professionalism would have to consist of a set of practices 'imbued with an ethics of integrative action that seeks to accommodate differing values and cultural outlooks'. Bottery (1998) covers some of the same ground, but provides a rather different account of what the embedded ethics might be: he suggests an appreciation of the provisionality of all knowledge and commitment to truth seeking, reflective integrity and humility and to client empowerment. These are 'the values which are needed to encourage the development of a "strong democracy" '.

4.2 Academic professionalism - pessimism or optimism?

Academics have been uneasy about the source of their professional identity; there has been a feeling that it ought to derive from a sense of unity with their peers in the academy but in practice it has derived from the professions for which they are preparing their students or from subject-based learned societies. In addition academics have become increasingly fragmented into a plurality of occupational groups, rewarded in different ways and with different contractual obligations (for example, academic-related staff and academic managers); the necessity to emphasise research output as the key indicator of quality has compounded the divisions.

One influential view of the academic profession is that between 1945 and the end of the 1960s it reached its high point ('when everything seemed possible in a benign state and a respectful society') and it declined thereafter (as it 'faced a hostile public led by stridently anti-academic politicians convinced of the truth of monetarism'). The quotations are from Halsey (1992). He famously explained the decline of donnish dominion in terms of proletarianisation: the power and advantage of the occupational group had suffered a three-fold reduction, principally in autonomy, but also in security of employment and chances of promotion.

The above suggests that academics are now living in a post-professional age. However, there are alternative readings. It may be that the academic profession has never really existed; in this sense academics are pre-professional. Or perhaps their profession has declined in relation to traditional criteria which are now themselves obsolescent; perhaps against different criteria there is an opportunity to claim a new professionalism. The professions will give way to professionalism? (Barnett, 2000 - p 157)

Both Piper (1994 - p 239) and Nixon (1997 b - p 100) argue that such a new professionalism can be constructed around the central idea of the academic as educator. This would require higher education to be managed in such a way that three things are encouraged: the recognition of teaching as an important area of professional expertise in its own right, the reintegration of teaching and research (in particular ensuring that teaching is research-based) and a commitment to a wide variety of research traditions (especially the core activity of research into teaching and learning within higher education). Elsewhere Nixon et al (1998) criticise those forms of professionalism which are inward-looking and self-interested and urge the emergence of a professionalism which is outward-looking and emancipatory. This would require an 'ethical turn' to a collegiality which forges supportive relationships against the grain of institutional management hierarchies.

These authors use the term ‘teaching’ throughout; their argument is, however, equally persuasive in terms of assessment. As with the Dearing Report, whenever teaching is mentioned we should explicitly add ‘and student assessment’. There is a significant advantage to be gained for academics by insisting that student assessment be seen as a set of professionally defined tasks. It may be unhelpful, for present purposes at least, to become too deeply immersed in whether or not academics are members of a profession; it may be more productive to focus on the opportunity which exists for student assessment to become recognised as a professional activity.

5. THE IMPACT OF STUDENT ASSESSMENT POLICY ON PROFESSIONALISM

Whether or not assessment practice is pre-professional, it is the case that assessment can be unprofessional. Unprofessional behaviour in relation to assessment is not rife within our universities; on the contrary, assessment is normally tackled with great conscientiousness and scrupulous attention to regulatory procedures. Nevertheless the proposition that unprofessional behaviour does occur on rare occasions is incontestable. There is no research evidence known to this writer which would allow general conclusions about the frequency of such behaviour; however, readers are asked to give thought to the items in Appendix A.

It is necessary to return now to the impact of emerging assessment policy on assessors (3.3) and to relate this to the characteristics of professionalism, old or new (4.1). In summary form, relationships are shown below.

Student assessment requires of assessors	Related professionalism
(i) content knowledge	(i) higher education - residual professionalism
(ii) craft knowledge	(ii) from experience - not directly related to professionalism
(iii) generic assessment scholarship	(iii) continuing professional education - emphasised in new professionalism
(iv) relinquishing of some autonomy	(iv) an attack on traditional professionalism
(v) increased range of ways of knowing, thus, need for team-working	(v) decrease in traditional, emphasis in new
(vi) consistency - in benchmarking - for increased reliability	(vi) requires cooperation and collegiality: new professionalism
(vii) complex discretionary nature of decision-making	(vii) sensitivity to moral dimensions: should characterise both old and new professionalism

(viii) critical thought and action

(viii) critically reflective practice:
central to new professionalism

(ix) training and continuing professional
development

(ix) a new professional body
required (a role for 'ILT?):
increase in traditional
professionalism and potential
for new.

Such a summary obscures the substance, subtlety and contestable nature of the links. What it does do, however, is suggest that the characteristics of traditional and new professionalism do not form unrelated matrices. There is not a simplistic dualism of contrasting forms from which one has to be chosen; rather the matrices require to be bisociated into some rather more creative redefinition of professionalism.

Only very guarded optimism about the development of any redefined professionalism seems justified: professionalism of any kind can only flourish if the environmental conditions are favourable. University teachers are being required not only to do different kinds of work, they are being required to do more and more work of all types. Without sufficient time for either critical reflection on their own practice or for cooperation with others, the likelihood of any growth in healthy professionalism is diminished; inadequate resources can themselves force a regression from the professional to the merely technical. When allied to an institutional culture which prioritises efficiency above fairness and justice to students, assessment procedures are even less likely to become more professional.

Freidson puts this as follows:

Without sufficient time, equipment, assistance and other resources, one cannot do one's work well and one's freedom to employ discretion becomes meaningless. If there are not enough resources available and the work must be done in any case, it is inevitable that it can only be done by the use of short-cuts and gross formulas that border upon mechanical and only marginally acceptable techniques. (Freidson, 1994 - p210)

The drama of student assessment is set to become more intense. The official scripts of policy may become more explicit, but they will continue inadequately to reveal the complexity of the impact of policy activity on the professionalism of academics.

¹ ILT is the Institute for Learning and Teaching in Higher Education, founded in 1999. Its first purpose is to enhance the status of teaching in higher education. Teaching presumably includes assessment. The first seminars of ILT have focused on assessment.

REFERENCES

- Barnett, R. (1997) *Higher Education: A Critical Business*, Buckingham: The Society for Research into Higher Education and Open University Press.
- Barnett, R. (2000) *Realizing the University - in an age of supercomplexity*, Buckingham: The Society for Research into Higher Education and Open University Press.
- Biggs, J. (1999) *Teaching for Quality Learning at University*, Buckingham: Society for Research into Higher Education and Open University Press.
- Bottery, M. (1998) *Professionals and Policy Management Strategy in a Competitive World*, London: Cassell.
- Bridges, P., Bourdillon, B., Collymore, D., Cooper, A., Fox, M., Haines, C., Turner, D., Woolf, H. and Yorke, M. (1999) 'Discipline-related Marking Behaviour Using Percentages: a potential cause of inequity in assessment', *Assessment and Evaluation in Higher Education*, Vol 24, No 3, pp285 - 300. (Referred to in Appendix A only.)
- Brown, S. and Glasner, A. (1999) *Assessment Matters in Higher Education*, Buckingham: Society for Research into Higher Education and Open University Press.
- Burchell, G. (1993) 'Liberal government and the techniques of the self', *Economy and Society*, Vol 22, No 3, pp 266-282.
- Colebatch, H.K. (1998) *Policy*, Buckingham: Open University Press.
- Freeman, R. and Lewis, R. (1998) *Planning and Implementing Assessment*, London: Kogan Page.
- Freidson, E. (1994) *Professionalism Reborn: Theory, Prophecy and Policy*, Cambridge: Eliot Freidson.
- Halsey, A.H. (1992) *The Decline of Donnish Dominion*, Oxford: Oxford University Press.
- HEQC (1996) *Guidelines on Quality Assurance (Section 7 - Student Assessment)*, London: Higher Education Quality Council.
- HEQC (1997) *Assessment in higher education and the role of 'graduateness'*, London: Higher Education Quality Council.
- NCIHE (1997) *Higher Education in the learning society: National Committee of Inquiry into Higher Education*, (The Dearing Report), London: NCIHE.
- Nixon, J., Martin, J., McKeown, P. and Ranson, S. (1997a) 'Towards a Learning Profession: changing codes of occupational practice within the new management of education', *British Journal of Sociology of Education*, Vol. 18, No 1, pp 5-28.

Nixon, J. (1997b) 'Regenerating Professionalism within the Academic Workplace' in Broadbent, J., Dietrich, M. and Roberts, J. (eds) *The End of the Professions? The Restructuring of Professional Work*, London: Routledge.

Nixon, J., Beattie, M., Challis, M. and Walker, M. (1998) 'What Does it Mean to be an Academic? A Colloquium', *Teaching in Higher Education*, Vol. 3, No 3, pp 277 - 298.

Ozga, J. (2000) *Policy research in educational settings: a contested terrain*, Buckingham: Open University Press.

Piper, D.W. (1994) *Are Professors Professional? The Organisation of University Examinations*, London: Jessica Kingsley Publishers.

QAA (1999) *Code of Practice for the Assurance of Academic Quality and Standards in Higher Education: External Examining - Draft Guidelines*, London: The Quality Assurance Agency for Higher Education.

Smith, H., Armstrong, M. and Brown, S. (eds) (1999) *Benchmarking and Threshold Standards in Higher Education*, London: Kogan Page.

Toohey, S (1999) *Designing Courses for Higher Education*, Buckingham: Society for Research into Higher Education and Open University Press.

Trowler, P.R. (1998) *Academics Responding to Change: New Higher Education Frameworks and Academic Cultures*, Buckingham: The Society for Research into Higher Education and Open University Press.

APPENDIX A

There follow seven anecdotes or vignettes. Readers are invited to consider each and provide their own answers to two questions. Are assessors professional? What kind of policy might help?

I “It does rather look as though my Council of Europe work will cut into the beginning of the autumn term, so I wonder if you’d be an angel and explain the basic principles of self-directed learning to my first year group, and if that doesn’t keep them quiet, offer a guaranteed B+ to the longest essay irrespective of merit.”

‘Postcard’ to Laurie Taylor Times Higher Education Supplement 27.8.93

II In a certain degree programme offered in the University of Glasgow an aggregated grade of 7 has great significance. At the Board of Examiners one student’s formula-derived grade was 6.9. There seemed to be an emerging consensus that this mark must be rounded up and the student given the benefit of any doubt. One senior lecturer held out against this and eventually won the day with this argument. ‘I well remember the first time I saw him coming up the stairs. I thought to myself then - there is a 6.9 student if ever I saw one. And nothing has happened since to convince me I was wrong. He is a 6.9.’

III James Scotland is a legendary figure in Scottish Education. He kept a copy of War and Peace on a reading stand beside the mirror in his bathroom so as not to waste time shaving. He lived a five minutes walk from his place of work and it was his regular practice to read and mark one student essay on his way home for lunch and another on the walk back. This practice was always taken as evidence of his ability to do at least two things at once. As far as can be recalled, no-one ever asked whether or not it was fair to the students.

IV “Dr X is just not to be trusted as far as assessment is concerned.”

“Why ever not?”

“ Well, two things really: first he’s a sexist and racist thug and secondly he always ensures that ‘his’ students are marked higher than anyone else’s.”

“He must have a problem with his black, female students...”

“Not really - word has got around, so such students never opt for his courses.”

Overheard in an English University - 1998

V In a keynote address to a conference on student assessment in Liverpool (November 1999), the speaker said: “I think we have probably all experienced that sick feeling when you leave a board of examiners knowing full well that some student has suffered a serious injustice. Is there anyone here who does not know that feeling?” One person raised his hand.

VI Student A takes joint honours History and English. Student B takes joint honours History and Computer Studies. In both instances the performance in History is average and the student is awarded 55%. In both instances the student performs very strongly in the second subject and achieves a mark which comes within the top 5% of the cohort. In English this mark is 75%, but in Computer Studies, where the assessor has confidence in recognising a near-perfect performance, the mark awarded is 95%. The outcome of this is that student A gains an overall mark of 65% (Upper Second Class) and Student B gains an overall mark of 75% (First Class). The performance of the two students has been closely comparable but one has the accolade of First Class while the other has to be content with the lesser award.

Bridges et al (1999 - p 285)

VII The present writer was at a social event recently where the conversation turned, without any prompting from him, to matters of professionalism.

A "I do feel as if all the old professions are letting the public down."

B "I do agree that medics are terribly arrogant."

A. "And lawyers are greedy. As for teachers..."

B. "Yes, always whingeing on about something or other."

CH. "What about academics?"

A "Academics? Academics? I have never thought of their being a profession at all."

Are assessors professional?

Student assessment and the professionalism of academics

COLIN HOLROYD *University of Glasgow and University of Stirling, UK*

ABSTRACT Student assessment is changing for educational reasons, in response to modularization and as a result of emerging policy. These changes have an impact on the professionalism of academics. Perhaps assessment was in the past a professional activity and has been proletarianized. In contrast, perhaps assessment was never genuinely professional and should be seen as being in a pre-professional phase. This article argues that assessment should have a key place in the regeneration of academic professionalism around the key role of the academic as educator. Improved assessment practice needs to draw upon a wider range of types of knowledge; it requires an institutional environment which encourages collegiality and communication; it should have, embedded within its codes, the values of equity, integrity and justice.

KEYWORDS: academic professionalism, assessment policy, higher education learning, student assessment

Introduction

The question in the title is intended to bring to mind the book 'Are Professors Professional?' (Piper, 1994). That book is undoubtedly an important one, but the title was surely chosen for its alliterative quality rather than its helpfulness. Its central concern is not with those who profess, but with those who assess, and more particularly with external examiners. It is the contention of this article that the issues are now even bigger and more important than those addressed by Piper. There are four reasons for this: (i) assessment is a more useful concept than examining and a much wider one, (ii) assessment is an important part of the role of all university teachers and not just professors, (iii) assessment is perceived more and more as a matter of central significance in higher education, and (iv) the professionalism of academics is increasingly questioned.

This article provides an exploration of some possible answers to four questions. How is student assessment changing? What does assessment

require of the assessors? How is the word 'professional' commonly used in relation to assessment? What is happening to academic professionalism?

How is student assessment changing?

General patterns of change

It is normal for texts on teaching, learning and assessment in higher education to give summary accounts of how assessment practices have changed over the last couple of decades. These accounts are not of course identical, but there is substantial overlap. To take just four of the most recent examples, important features of change are identified by Biggs (1999a; chapter 8), Brown and Glasner (1999; chapter 2), Freeman and Lewis (1998; p. 310) and Toohey (1999; chapter 9).

Important dimensions of change are said to include the following:

- (i) increasing emphasis on the learning enhancement purpose of assessment rather than its certification and accountability purposes;
- (ii) increased attention to formative rather than summative aspects;
- (iii) more emphasis on a standards model of assessment, involving criterion-referenced assessment, and less on a measurement model, involving norm-referenced assessment;
- (iv) more frequent provision of descriptive comment and constructive feedback and less restriction of assessor response to marks, grades and summary labels;
- (v) a move from dependence on one main method of assessment (and end-of-course assessment) to deploying a variety of methods (and within-course assessment);
- (vi) less reliance on assessment by teaching staff alone and more involvement of self, peers and workplace assessors;
- (vii) increased insistence on assessment as integral to teaching rather than a separate activity occurring after teaching.

Of course, any such list immediately prompts questions and debate; these cannot be pursued at any length here. It should perhaps be noted, however, that these dimensions of change are not separate and distinct; several of them are closely related and item (iii) is at a higher level of generality and includes some of the others. Also, to identify these as system-wide trends has to allow that there will be local contravention of such trends; there are, no doubt, many places and courses where change has been in an opposite direction. There are also some pervasive countervailing pressures which could result in significant reversals. For example, it is frequently claimed that validity is in the ascendant in assessment thinking

and activity; this is evidenced by the move to more authentic forms of assessment and the insistence that assessment procedures be aligned more constructively with statements of intended learning outcomes (Biggs, 1999b). However, others would argue that pressures for accountability, transparency and consistency from government and from potentially litigious students-as-consumers mean that reliability is a more compelling requirement than ever before.

Change related to modularization

There is a tendency within the literature for it to be assumed that all the important decisions about assessment are to do with selecting the method of assessment and then marking the products of individual students. This leaves out a great deal, as will be shown in the next section; crucially it omits the management of assessment and decisions on aggregation and final awards. This latter decision-making phase has been altered profoundly by relatively recent moves towards the modularization of higher education courses and programmes. The impact of these is helpfully described and discussed by Wolf and her co-authors in the report *Assessment in Higher Education and The Role of 'Graduateness'* (Higher Education Quality Council [HEQC], 1997).

Wolf distinguishes two ways of modelling how universities assess students' work and make decisions about degree awards. In the first, the bulk of assessment is by end-of-year examination. There may be some options, but the course is not part of a full-scale modular system. The examiners (both internal and external) know each other well and most have considerable expertise in examining this kind of degree. Examination scripts are usually double-marked blind. There will probably have been a meeting of the examining board to scrutinize questions, marking criteria and mark schemes. All examiners attend the board and final decisions on awards and classes are made there. There is the opportunity for shared understandings about assessment criteria and standards to develop. This 'type' has not yet disappeared at undergraduate level, but it applies to a decreasing proportion of undergraduates. Many sub-sections of degree programmes approximate to it and it is still typical at postgraduate level. Twenty years ago it would have been far more common and it is the approach into which most academics were themselves socialized.

The second model typifies the awarding process for modular degrees in which many subjects are taken as options and in which many 'home programme' students have selected options from a range of other programmes. Not all work has been double-marked. Written criteria may not be drawn up by the programme teams. There have been no scrutiny meetings. For any one given student, many of the staff making assessment decisions will

have had little experience of examining together. Only a small number of examiners attend the examination boards; final decisions may well be taken later at programme and faculty boards. This second type encapsulates trends in higher education which are recent, but which currently show no signs of reversing themselves.

The emergence of this second model is hugely significant. The research evidence points to the great importance of assessors being members of a communicating network; only in this way do similar meanings get attached to the standards implicit in assessment criteria (e.g. Wolf, 1995; p. 77). When few examiners are involved in programme or degree-wide decision-making fora, then their socialization has been into other groups. They (and external examiners) will necessarily be presented with assessments from subject areas in which their own expertise is weak and with issues of aggregation which are technically insoluble. Under these circumstances their predominant concerns will have to be for procedural correctness and 'due process' and their decisions will be largely formula-driven. These features clearly relate to the professionalism of assessors.

Policy-related changes

Student assessment is changing, and will continue to change, because of the recent and continuing emergence of national policy on assessment. Although it would not be true to suggest that such policy has only been detectable in the last 2 or 3 years, it is only since the report of the National Committee of Inquiry into Higher Education (NCIHE, 1997) – the Dearing Report – that this has been an obvious feature. The NCIHE made recommendations to government and, of course, recommendations do not automatically become official national policy, as the issue of student fees makes very clear. Nevertheless it can be predicted with confidence that the report as a whole will have a significant influence on the general thrust of policy development including that on assessment. Developments are ongoing, so what follows is necessarily tentative and speculative.

There is a seeming paradox. The Dearing report attaches a high priority to the promotion of learning and the enhancing of standards of student achievement. Why then in its 88 main recommendations is there no explicit mention of student assessment at all? (The word 'assessment' is only used in relation to the research assessment exercise in recommendations 32, 33 and 34.) The full report (NCIHE, 1997, Vol. 1) does make explicit reference to student assessment and addresses the topic directly on pages 137–39, but oddly this section results in no recommendations. Does this demolish any claim that the Dearing report will significantly affect assessment practice? The paradox disappears if we decide that Dearing saw assessment as a part of any effective strategy for the

management of learning – perhaps one that seemed so obvious that explicit reference was judged unnecessary.

If this view of assessment is taken, then several recommendations have significant implications for student assessment. Noteworthy are: the priority to be attached to promoting student learning (recommendation 8); the training and accreditation of university teachers (recommendations 13 and 48); the development of programme specifications (recommendation 22); the amendment of the remit of the Quality Assurance Agency (QAA) to include standards verification and the development of a code of practice (including a section on external examining and student assessment) to be adopted by all institutions as a condition of continued public funding (recommendation 24) and the setting-up of small expert teams to provide benchmark standards/statements (recommendation 25). Several of these topics deserve more extended discussion and comment than is possible here [see, for example, Smith, Armstrong and Brown (1999)]. Some quick comments, however, are required on benchmarking and on the code of practice in relation to external examiners and student assessment.

Benchmark statements will define for the UK as a whole, and for the first time, what has to be assessed within honours degrees in 42 different subject areas. It is important to ask just how far the prescriptiveness of these benchmark statements will extend. First, the statements apply to honours degrees within a discipline, but they will have a major impact on all those courses through which students progress to honours level and then indirectly on the combinations of these courses into multidisciplinary degrees. Second, it is made wholly explicit that statements are not to define subject content – this is to remain a matter for individual institutions. There does seem to be just a whiff of disingenuousness here; it is not obvious that understanding within a discipline (which is to be included in benchmark statements) can exist independently of the subject content which is understood (which is not included in benchmarking). Third, the statements prescribe what is to be assessed, but not how it is to be assessed; however, such a prescription taken along with the exhortation that assessment procedures be valid does exert a strong influence on methods of assessment.

Recommendation 24 of Dearing concerns the development of a suite of inter-related documents which together will form a code of practice which will be adopted by all universities as a condition of continued funding. The section of this code devoted to external examining has already been published (QAA, 1999). Rather than looking at the substance of the document (which is unsurprising) it is instructive for present purposes to consider its form. The document articulates a series of 19 precepts each of which has accompanying outline guidance. The precepts 'identify those key matters which the Agency expects an institution to be able to demonstrate

it is addressing effectively through its own quality assurance mechanisms'. The guidance is meant to be neither prescriptive nor exhaustive, but to offer a framework which individual institutions may wish to adopt or adapt to meet their own cultures; 'nevertheless, in many institutions the guidance will constitute appropriate good practice'. If a code is a set of rules or a body of laws, if a precept is a command, rule or instruction and if guidance defines what should 'normally' be seen as good practice, then it seems we do have here a national policy which it will be difficult for any university to ignore.

The document within the code dealing with student assessment in general is expected to be published in the second half of 2000. If publication is to be as early as this (as the government is said to wish), then the substance is likely to be very similar to that of the *Guidelines on Student Assessment* produced by the QAA's predecessor, the Higher Education Quality Council (HEQC, 1996). It seems likely that the form will be the same as that of the section on external examining, i.e. a set of precepts with accompanying guidance as to what normally constitutes good practice.

Given that policy is only beginning to emerge, what is likely to be the essence of that policy with regard to student assessment?

- (i) Programme specifications must show that intended learning outcomes and the assessment procedures to show the level of student achievement of these are positively aligned.
- (ii) Nationwide expectations about the comparability of standards and the quality assessment procedures in higher education will need to be met; institutions must demonstrate that their locally determined assessment procedures effectively address the nationally defined precepts.
- (iii) Both internal and external examiners will need to receive such training, support and continuing professional development as is needed for their activities to meet the necessary quality standards.

What does assessment require of the assessors?

Aspects of assessment

It is common for academics to ask what their assessment procedures require of those assessed, i.e. their students; it is rare to ask what assessment requires of those who do it, i.e. the assessors.

There is always a risk of discussing student assessment as if it were all one activity, or perhaps a set of activities all roughly similar in type. Student assessment is a complex business which can involve, this writer would

argue, at least 11 distinct types of activity: (i) the design and management of assessment systems; (ii) the creation of constructed-response assessment instruments; (iii) the grading of student responses to these and the provision of feedback; (iv) the creation of objective and computer-managed test-items; (v) the marking of student answers to these; (vi) the creation of competence-based assessment procedures; (vii) the assessment of student performance defined in competence terms; (viii) the development and implementation of self and peer assessment; (ix) the monitoring and modification of assessment performed by others; (x) that informal assessment which is integral to teaching; (xi) the decision-making procedures in boards of examiners.

Requirements of assessors

Our original question should now be modified to, 'What does each of the above aspects of assessment require of assessors?' A full answer would need to deal with: (a) the different types of knowledge on which assessors should draw; (b) the range of types of decision-making involved; (c) the extent to which any one assessor can work alone or needs to collaborate with others; and (d) understanding of relevant national and institutional policy constraints. This full answer cannot be provided here, but outline answers for four or five aspects should illustrate some important general issues.

The creation of constructed-response instruments clearly requires subject content knowledge; most academics have substantial 'craft' knowledge of this assessment activity (from their own experience of being assessed and assessing others); any training needed is likely to be minimal; there is some scope for what might be called scholarly knowledge of assessment, but assessors are usually likely to 'get by' without it (although they may not avoid assessment bias). The decisions required are relatively simple, but the assessor has a fair amount of room for discretionary judgement within the course or programme specifications. This activity can be carried out by the solitary assessor, indeed currently it sometimes is; but the benefit of having questions criticized by colleagues has long been accepted. In the future there will be a much greater need for communication between assessors on this activity because of currently emerging policy, for example, to ensure that individual assessment items have face validity in the light of some shared understanding as to what the benchmark statements for a particular subject mean.

Grading and provision of feedback in assessment of this type requires subject content knowledge and assessment craft knowledge; there is some scope for training and the acquisition of a little assessment scholarship. The provision of constructive feedback would be enhanced by greater

educational knowledge. However, the overwhelming need here is for the recognition that consistency in grading only comes about when assessors, again, are members of a communicating group. The notion that consistency problems in assessment are solved by the production of a set of assessment criteria is woefully simplistic; there is a pressing need for the development of shared understanding of what the criteria mean in practice. Standards are established through argument, discussion and interaction over time. Policy demands in relation to standards will only be satisfied if assessors have the opportunity to communicate, cooperate and collaborate.

The creation of selected-response items, point (iv) above, makes similar demands on subject content knowledge; assessment craft knowledge, however, does not take people very far – it usually results in badly constructed multiple-choice tests of simple factual knowledge; there is much greater scope for training and for assessment scholarship; computer-based assessment obviously makes new demands on knowledge of information technology. Assessment scholarship will help avoid class, sex and gender bias. The assessor still has considerable scope for discretion in decision-making and the necessary decisions are more complex. This activity has never been effectively carried out by a lone assessor: the benefits of working in a team have usually been recognized. Technological advances increase the need for communication within collaborating groups, as do the implications of national benchmarking policy. However, marking these items, point (v) above, carries no knowledge requirement – it is formulaic; decisions are simple and non-discretionary; the activity can be solitary.

Aggregation decisions and award-making, point (xi) above, call on a huge range of knowledge bases (this does not mean that each person is expected to have expertise in all areas): these include, but are not limited to, subject knowledge, assessment craft knowledge, assessment and educational scholarship, information technology knowledge, national and institutional policy, management knowledge, interpersonal knowledge. Decision-making is (or should be) hyper-complex and discretionary. The pressure from modular systems to increase the formulaic nature of examining boards has already been noted; in the interests of justice to students, assessment boards must not become the sole preserve of managers and administrators.

This type of analysis could be continued for the remaining aspects of assessment. Although clearly much remains contentious, three general aspects are worthy of comment.

It would be foolish to underestimate the importance of discipline-specific content knowledge and assessment craft knowledge. It is essential for assessors to know their subject well, certainly to have a great deal more

understanding of the subject than is likely to exist in those assessed. Assessment of a subject deserves scholarship in that subject. By assessment craft knowledge is meant that understanding of assessment that inevitably comes from experience of assessing and some, however minimal, reflection upon it. Craft knowledge is of course implicit and even experienced university teachers can have difficulty in making their understanding explicit to others; without the pressures to explication which come from assessors choosing, or being required, to communicate with each other there is a risk of unexamined beliefs persisting.

However, to recognize that subject expertise is necessary and that some craft knowledge is inevitable is not to accept that taken together they are sufficient. Or rather they may have been sufficient to get by in the past – but they are no longer sufficient to meet the demands of high-quality assessment for a student population much bigger, more diverse and more critical than in the past.

A distinction is necessary between assessment craft knowledge and assessment scholarship. Without opening a debate as to whether assessment can be regarded as a scholarly discipline in its own right, it is indisputable that there is a substantial body of research and literature on assessment in higher education which at least can help assessors escape the obvious pitfalls and which at best can illuminate better practice. Two quick examples must suffice.

First, more than a few university teachers have somewhat hastily introduced peer assessment into their courses on a vague intuition that this is a procedure whose time has come. They are then surprised when students report some reluctance to become involved and then seem either unjustifiably harsh in their judgements of their colleagues or sentimentally overgenerous. Even a very cursory browse through the recent research would convincingly demonstrate the crucial importance of students being carefully inducted into the process. Second, in a recent 'expert-led' seminar on computer-based assessment, the speaker assumed that multiple-choice questions were the obvious item format for objective tests and assumed that the best number of options to provide within each item was five. Even a modest acquaintance with research literature from about 20 years ago would have been helpful.

It could of course be argued that there is now so much to be known about assessment that it is unreasonable to expect any hard-pressed university teacher to know it all. This might then lead to the proposal that there should be a separate group of assessment experts, in addition to and detached from, university teachers. This would be regrettable. Assessment is an essential element of teaching and only separable from it with serious consequences; it is too important to be left to 'experts'. A more satisfactory

conclusion, based on acceptance that university teachers do need to know more about assessment and that different teachers have different knowledge strengths, is that team-working, communication, collaboration and cooperation are essential.

Assessment does not have one specialized knowledge base – it draws on many types of knowledge. Decision-making in most, but not all, aspects of assessment is difficult, complex and discretionary. Assessment is a matter of judgement based on inferences from evidence evaluated as valid, reliable and fair; it is crucially concerned with justice for students in its award-making aspects and with the empowerment of students in its learning enhancement aspects. These are the central justifications for regarding assessment as a professional activity in its own right.

Emerging national policy in assessment is something of a double-edged sword in relation to professionalism. It seriously restricts the autonomy of university teachers, i.e. it reduces their professionalism as traditionally defined. However, it also increases the necessity for assessors to communicate effectively; there is an opportunity here for a new professionalism giving a central place to collegiality.

Two common usages of 'professional' in relation to assessment

'Assessment should remain a matter of professional judgement'

In universities around the country there are groups of people working away to produce the institution-specific documents on assessment required by the QAA. In one of these familiar to this writer, the first sentence in the draft policy is 'Assessment is a professional academic judgement about the extent to which a candidate has achieved the intended learning outcomes'. Informal contact suggests that something similar is usual elsewhere. Why are such statements given prominence? There seem to be three related reasons.

First, there is a strong wish to emphasize that assessment decisions should always go beyond the technical and formulaic. Objective testing may permit a mechanical computation of marks and scores, but the attachment of meaning and significance to these is inescapably a judgement to be performed by persons. There is a closely related distrust of competence-based assessment in which the definition of desired competences, along with range statements and some prescription of underpinning knowledge instrumentally conceived, all lead to assessment which is seen as atomistic, reductionist and technicist – the much-reviled 'tick-the-box' approach.

Second, there is an awareness that assessment by its nature must be an uncertain and problematic activity. It always requires inferences to be made about the internal mental states of complex human beings from more or less inadequate evidence provided in samples of observable behaviour and artefacts/products. Third, there is the fear that pressures from increased numbers of students will result in the delegation of too many assessment activities to administrative staff, to graduate teaching assistants and to computers for economic reasons alone. The stress on assessment as a professional activity is a plea for the important decisions in assessment to remain the preserve of academic staff. Academic staff do not enjoy the burden that assessment increasingly places upon them, but they do not wish to hand it over to others.

'The assessor's behaviour was not professional'

A useful distinction has been drawn between 'being a professional' and 'behaving as a professional' (Helsby, Knight, McCulloch, Sounders and Warburton, 1997). Whether or not it is agreed that the academic assessor belongs to a profession, it can be agreed that certain behaviours should be criticized as unprofessional. This is not to assert that unprofessional behaviour in relation to assessment is rife within our universities; on the contrary, assessment seems normally to be tackled with great conscientiousness and scrupulous attention to regulatory procedures. Nevertheless, the proposition that unprofessional behaviour does on occasions occur seems incontestable. There is no research evidence available which would allow general conclusions about the incidence of such behaviour. However, readers are asked to reflect on one anecdote and ask themselves what memories it prompts.

At an assessment conference in Liverpool in November 1999, a key speaker posed this question to an audience of about 50 academics. 'You know that sick feeling you have when you leave a board of examiners convinced that something has gone wrong? That some student has been unfairly treated? Is there anyone here who has never had that feeling?' There was – one person.

On rare occasions, assessors behave in unprofessional ways. Any one occurrence is of great seriousness for the student(s) concerned. Are we certain the necessary codes are in place to ensure the absolute minimum of unprofessional behaviour? And to ensure appropriate redress for those students who suffer from any occurrence?

What is happening to academic professionalism?

Changing characteristics of professionalism

Some analysis of professionalism seems inescapable. The problem, of course, is that professionalism is not some social-scientific absolute, but a historically changing and socially constructed concept-in-use. The extensive literature variously describes professionalism as traditional, quintessential or residual; as semi-, restricted or extended; as dispersed, hybrid or syncretistic; as corporatist or bureaucratic; as re-imagined, redefined, retheorized, reconceptualized, regenerated, reborn or reclaimed.

There is some measure of agreement that traditional public-sector professionalism had four defining features:

- (i) Professionalism was made manifest in the pursuit of an occupation or calling which provided the person with a living (as distinct from amateurism).
- (ii) The members of a profession had completed both some form of higher education which involved the mastery of some difficult body of useful knowledge and some extended form of training in relevant skills.
- (iii) The primary allegiance was to a self-controlling professional body rather than to an employer; this body acted as guardians of the specialist knowledge, gatekeepers to membership and enforcers of an ethical code.
- (iv) The substantial gap between the knowledge and competence of the professional and of the client and manager meant it was difficult for these others to evaluate the competence of the professional and thus professionalism necessarily implied a degree of autonomy, of self-control and of trust.

Freidson (1994) effectively summarizes all this into two basic requirements of professionalism: first, the taking of complex, discretionary decisions should be followed by effective action (i.e. the requirement to do good work) and second, that the work should have a clear value for the well-being of society as a whole (i.e. the requirement to do Good Work).

Traditional professionalism entered a period of crisis in the 1980s. A number of reasons are offered for this: the debasement of some professions into self-serving cabals; the proliferation of groups making flimsy claims to status for financial gain; the fragmentation of existing professional groups; growing public distrust of experts; growing managerialism, accountability and state control and a retheorizing of the status of specialized knowledge. The period since has witnessed the evolution of new forms of professionalism which may be better adapted for survival in

a post-modern environment. For example, Nixon et al. (1997) see a regenerated professionalism as necessarily involving collegiality (among the members of the profession), negotiation (between professionals and clients), cooperation (with members of other professions) and partnership (between professionals and other stakeholders). This new professionalism, he argues, will have to consist of a set of practices 'imbued with an ethics of integrative action that seeks to accommodate differing values and cultural outlooks'. Bottery (1998) covers some of the same ground, but provides a rather different account of what these embedded ethics might be: he suggests an appreciation of the provisionality of all knowledge and a commitment to truth seeking, reflective integrity and humility, and client empowerment. These are 'the values which are needed to encourage the development of a "strong democracy" '.

Academic professionalism – a pessimistic view

Academics have been uneasy about the source of their professional identity; there has been a feeling that it ought to derive from a sense of unity with their peers in the academy, but in practice it has derived from the professions for which they are preparing their students or from some other external source (for example, the Royal Society of Chemistry or the British Psychological Society). In addition, academics have become increasingly fragmented into a plurality of occupational groups, rewarded in different ways and perhaps even with different contractual obligations (for example, academic-related staff and academic managers); the necessity to emphasize research output as the key indicator of quality has compounded the divisions.

One influential view of the academic profession is that between 1945 and the end of the 1960s it reached its high point ('when everything seemed possible in a benign state and a respectful society') and it declined thereafter (as it 'faced a hostile public led by stridently anti-academic politicians convinced of the truth of monetarism'). The quotations are from Halsey (1992). He famously explained the decline of donnish dominion in terms of proletarianization: the power and advantage of the occupational group had suffered a three-fold reduction, in autonomy, in security of employment and in chances of promotion. There are, however, two features of Halsey's magisterial work which are slightly surprising: he provides no explicit analysis of what it means to belong to a profession and no sustained scrutiny of the claim from academics to be professional. If this seems an impudent comment, consider three things. First, Halsey oscillates between describing academics as one profession ('the key profession') and several ('What has happened to the academic professions?'). Second, he uses the telling phrase: '... and the professional classes to which it (i.e. higher

education) has added itself and which it serves'. In other words he sees the claim to professionalism as based on its serving other professions rather than on itself possessing the defining characteristics of professionalism. Third, the key profession has been proletarianized and as a result 'dons are no longer set apart so clearly from the professional, administrative and industrial classes' (Halsey, 1992; p. 146). Academics have, it seems, declined from being the servants of the professions to being more closely identified with them; there must surely be some irony intended.

Academic professionalism – is there any more optimistic view?

The above suggests that academics are now living in a post-professional age. Of course, there are alternative readings. It may be that academic professionalism has never genuinely existed; in this sense academics are still pre-professional. Or perhaps their professionalism has declined in relation to criteria which are no longer the appropriate ones, and against different criteria there is now an opportunity to claim a new professionalism.

Both Piper (1994: 239) and Nixon (1997: 100) argue that such a new professionalism can be constructed around the central idea of the academic as educator. This would require higher education to be managed in such a way that three things are encouraged: the recognition of teaching as an important area of professional expertise in its own right, the reintegration of teaching and research (in particular ensuring that teaching is research-based) and a commitment to a wide variety of research traditions (especially the core activity of research into teaching and learning within higher education). Elsewhere Nixon et al. (1998) criticize those forms of professionalism which are inward-looking and self-interested and urge the emergence of a professionalism which is outward-looking and emancipatory. This would require an 'ethical turn' to a collegiality which forges supportive relationships against the grain of institutional management hierarchies.

These authors use the term 'teaching' throughout; their argument is, however, equally persuasive in terms of assessment. As should be done for the Dearing report, whenever teaching is mentioned we should openly add 'and student assessment'. There is a significant advantage to be gained for academics by insisting that student assessment be seen as a set of professionally defined tasks.

Academic identity continues to be informed by some pre-professional beliefs and myths about assessment which appear increasingly problematic in this time of relentless change. Taylor (1999: 138) concludes that such beliefs limit academics' capacities to generate constructive, considered and collegial responses in the face of pressures that currently threaten the very concept of 'academic'. These limiting beliefs are ripe for

replacement. Such a process would require clear recognition that student assessment involves:

- a complex set of activities informed not by one specialized knowledge base but by a range of different types of knowledge
- a move beyond the necessary content knowledge and assessment craft knowledge into assessment scholarship – which will involve continuing training and development support
- a range of types of decision-making, most of which remain difficult and discretionary even when national policy means some reduction in autonomy
- effective interaction with all the assessment stakeholders requiring collegiality, communication, cooperation, partnership and negotiation
- a positive emphasis on that assessment which empowers students through enhanced learning
- a code of practice in which are firmly embedded values of equity (as seen in the removal of assessment bias), justice (in both the awarding of grades and in their aggregation), truth-seeking and an insistence on personal responsibility (to combat student dishonesty)
- a commitment to critical reflection on personal assessment practice and to research on assessment whether personal, subject-based or institutional.

These seven features can give a basis for professionalism in assessment activity. They should not be seen as providing a justification for creating a new expert group within higher education: 'the assessment professionals'. This is because assessment is intimately intertwined with teaching and learning and should not be detached from them; if the new professionalism of academics is to be centred on their role as educators, then that role has to retain assessment at its core. Assessment is something which academic educators must do; it cannot be hived off to some separate specialist group.

Only very guarded optimism about the emergence of this new professionalism seems justified. Professionalism of any kind can flourish only if the environmental conditions are favourable. University teachers are being required not only to do different types of work, they are being required to do more work of all types. Without sufficient time for either critical reflection on their own practice or for cooperation with others, the likelihood of any growth in healthy professionalism is diminished; inadequate resources can themselves force a regression from the professional to the merely technical. When allied to an institutional culture which prioritizes efficiency above fairness, assessment procedures are even less likely to become more professional. Freidson (1994) puts this as follows.

Without sufficient time, equipment, assistance and other resources, one cannot do one's work well and one's freedom to employ discretion becomes meaningless. If there are not enough resources available and the work must be done in any case, it is inevitable that it can only be done by the use of short-cuts and gross formulas that border upon mechanical and only marginally acceptable techniques. (p. 210)

The opportunity exists for assessors to become more professional. The single most effective way of enhancing learning within higher education is through the improvement of assessment procedures. Assessment is at the core of the academic role of educator and on that role can be centred a reclaimed professionalism for academics. However, professionalism as educators – including professionalism in assessment practice – does not come cheap; commitment to a policy of better learning through higher quality assessment will have to be supported by an increased allocation of resources.

References

- BIGGS, J. (1999a) *Teaching for Quality Learning at University*. Buckingham: Society for Research into Higher Education and Open University Press.
- BIGGS, J. (1999b) 'What the Student Does: Teaching for Enhanced Learning', *Higher Education Research and Development* 18(1): 57-75.
- BOTTERY, M. (1998) *Professionals and Policy Management Strategy in a Competitive World*. London: Cassell.
- BROWN, S. & GLASNER, A. (1999) *Assessment Matters in Higher Education*. Buckingham: Society for Research into Higher Education and Open University Press.
- FREEMAN, R. & LEWIS, R. (1998) *Planning and Implementing Assessment*. London: Kogan Page.
- FREIDSON, E. (1994) *Professionalism Reborn: Theory, Prophecy and Policy*. Cambridge: Eliot Freidson.
- HALSEY, A.H. (1992) *The Decline of Donnish Dominion*. Oxford: Oxford University Press.
- HELSEBY, G., KNIGHT, P., MCCULLOCH, G., SOUNDERS, M. & WARBURTON, T. (1997) *Professionalism in Crisis: A Report to Participants on the Professional Culture of Teachers Research Project*. Lancaster: University of Lancaster.
- HEQC (1996) *Guidelines on Quality Assurance (Section 7 – Student Assessment)*. London: Higher Education Quality Council.
- HEQC (1997) *Assessment in Higher Education and the Role of 'Graduatness'*. London: Higher Education Quality Council.
- NCIHE (1997) *Higher Education in the Learning Society: National Committee of Inquiry into Higher Education. (The Dearing Report.)* London: NCIHE.
- NIXON, J. (1997) Regenerating Professionalism within the Academic Workplace in J. Broadbent, M. Dietrich and J. Roberts (eds) *The End of the Professions? The Restructuring of Professional Work*. London: Routledge.
- NIXON, J., BEATTIE, M., CHALLIS, M. & WALKER, M. (1998) 'What Does it Mean to be an Academic? A Colloquium,' *Teaching in Higher Education* 3(3): 277-98.
- NIXON, J., MARTIN, J., MCKEOWN, P. & RANSON, S. (1997) 'Towards a

- Learning Profession: Changing Codes of Occupational Practice Within the New Management of Education', *British Journal of Sociology of Education* 18(1): 5–28.
- PIPER, D.W. (1994) *Are Professors Professional? The Organisation of University Examinations*. London: Jessica Kingsley.
- QAA (1999) *Code of Practice for the Assurance of Academic Quality and Standards in Higher Education: External Exemining – Draft Guidelines*. London: The Quality Assurance Agency for Higher Education.
- SMITH, H., ARMSTRONG, M. & BROWN, S. (eds) (1999) *Benchmarking and Threshold Standards in Higher Education*. London: Kogan Page.
- TAYLOR, P.G. (1999) *Making Sense of Academic Life: Academics, Universities and Change*. Buckingham: Open University Press.
- TOOHEY, S. (1999) *Designing Courses for Higher Education*. Buckingham: Society for Research into Higher Education and Open University Press.
- WOLF, A. (1995) *Competence-based Assessment*. Buckingham: Open University Press.

Biographical note

COLIN HOLROYD is an honorary senior research fellow at the University of Glasgow and a research student at the University of Stirling. His current research interests lie in assessment policy and assessment practice in higher education. He would welcome comments on any aspect of this paper.

Address: Teaching and Learning Service, University of Glasgow, Glasgow G12 8QQ, Scotland. [email: c.holroyd@admin.gla.ac.uk]

APPENDIX XI

RESEARCH SEMINAR: TASKS AND RESPONSES

There were 26 participants divided into six groups, each of four or five people. There were three tasks; two groups tackled each of the three tasks.

Task A: (Appoint a scribe. As a group, explore your answers to these questions. Summarise the main points for reporting back.)

Have changes in how you assess your students affected your academic professionalism? If so, how?

Task B: Have changes in University assessment policy affected your academic professionalism? If so, how?

Task C: Have changes in assessment policy from external sources affected your academic professionalism? If so, how?

Participants engaged enthusiastically with the tasks and appeared to think the questions worth asking and trying to answer. However, responses were remarkably 'thin'. The significant points were as follows.

- You asked us to do something unfamiliar; we needed longer to think about the tricky issues involved.
- The questions do matter, but we find our own answers unsatisfying.
- We all seemed to think about professionalism differently; we needed to sort out our meanings first to make more progress.
- Is professionalism the most useful concept? Why not just academic freedom and autonomy?
- Critical professionalism should activate us to resist interference.
- Policy is a weapon used by central managers to change how we assess students.
- Policy usually seems to restrict professionalism
- Policy from the University, and from outside sources, never seems to take subject differences properly into account.
- Boards of examiners used to be meetings of professionals about people; now they are about aggregation formulae.
- Central control and policy make it difficult to hold on to professionalism.
- Our professionalism should mean we protest more about imposed change, but do we?
- As conditions get increasingly difficult, we are going to become increasingly unprofessional in how we assess our students.
- External policy will increasingly change assessment in future.

APPENDIX XII

ASPECTS OF APPROVAL AND ANONYMITY

Proposed research on methods of assessment

1. The purposes of the research are all related to description, understanding and illumination; there is no purpose related to evaluation or criticism. The focus is exclusively on the rationales for different assessment methods.
2. In any report, oral or written, arising from the research, the following rules will be observed.
 - (i) No individual member of staff will be named.
 - (ii) Subject areas will be named.

The research will only take place in departments where the head of department agrees to the naming of the subject area.

(iii) The University of Glasgow will not be named in any published paper or conference presentation.

Although The University will not be named, it should be noted that it may be identifiable through a determined study of, say, acknowledgements and references.

3. The above rules will allow statements of the following types to be made.

Five of the eight members of staff working in this subject area said that change had occurred because....

'This was an attempt to reduce the assessment burden on both staff and students.....'
(Staff member 6)

Agreement to seek interviews with seven members of staff of the chemistry department was obtained from the head of that department.

The interviews took place in four departments of a university in the United Kingdom; it is referred to from now on as The University.

4. All people with whom interviews are sought will be sent a copy of this protocol and a letter which outlines
 - (a) the purposes of the research
 - (b) the main topics to be covered in interview
 - (c) an estimate of the probable length of the interview.
5. After all the interviews within a department have been completed, a draft of a report giving findings from that subject/department will be sent to each person interviewed (and to the head of department if s/he was not an interviewee), providing an opportunity for him or her
 - (a) to correct errors of fact
 - (b) to indicate things s/he wishes to be omitted.

A summary of the final version of this report will be sent to the relevant head of department for distribution as s/he believes appropriate.

INTERVIEW SCHEDULE

SUBJECT AREA: DESIGN

Points in preamble

- (a) Thanks for your time.
- (b) The research purpose is..... NB Rationale for practice and change.
- (c) Can I have your permission to record this? Only to make life easier for me.
- (d) How long this takes depends on how much you want to say. Any deadline?
- (e) Anything from you before we plunge in?

Can we start please by focusing on you? You are AB and your title is?

How long have you had any involvement in the assessment of design students?

Thinking about the current design curriculum, what roles do you have in relation to the assessment of students?

In talking about assessment shall we focus on any particular year or years – or shall we try to deal with all four years?

- A. What in your view are the good features in how students are assessed within the design curriculum?
- B. What seem to you to have been the most significant changes in assessment practice in design in recent years?
- C. In your view, why were these changes introduced?

Have these changes in assessment practice had any significant impact on your work as an academic? Or on the work of your colleagues?

- D. If the conditions could be made favourable, are there changes in current assessment practice that you personally would like to see? What are they?

Why would you like to see these changes?
What are the conditions necessary to make these changes successful?

- E. Do you think you may be required to make changes in how students are assessed in the near future? What changes?

Would you personally welcome these changes?

What are the reasons that make it likely that these changes will be required?

How will such changes impact on the work of academics?

F. You have given me your reasons for changes in how design students are assessed. I'd very much like to have your reaction to some of the reasons that other people give.

Sheet 1 The main purposes of assessment.

Sheet 2 Reasons for change already made

Which of these will become more influential in future?

G. How important is assessment?

Does it take up too much of your time?

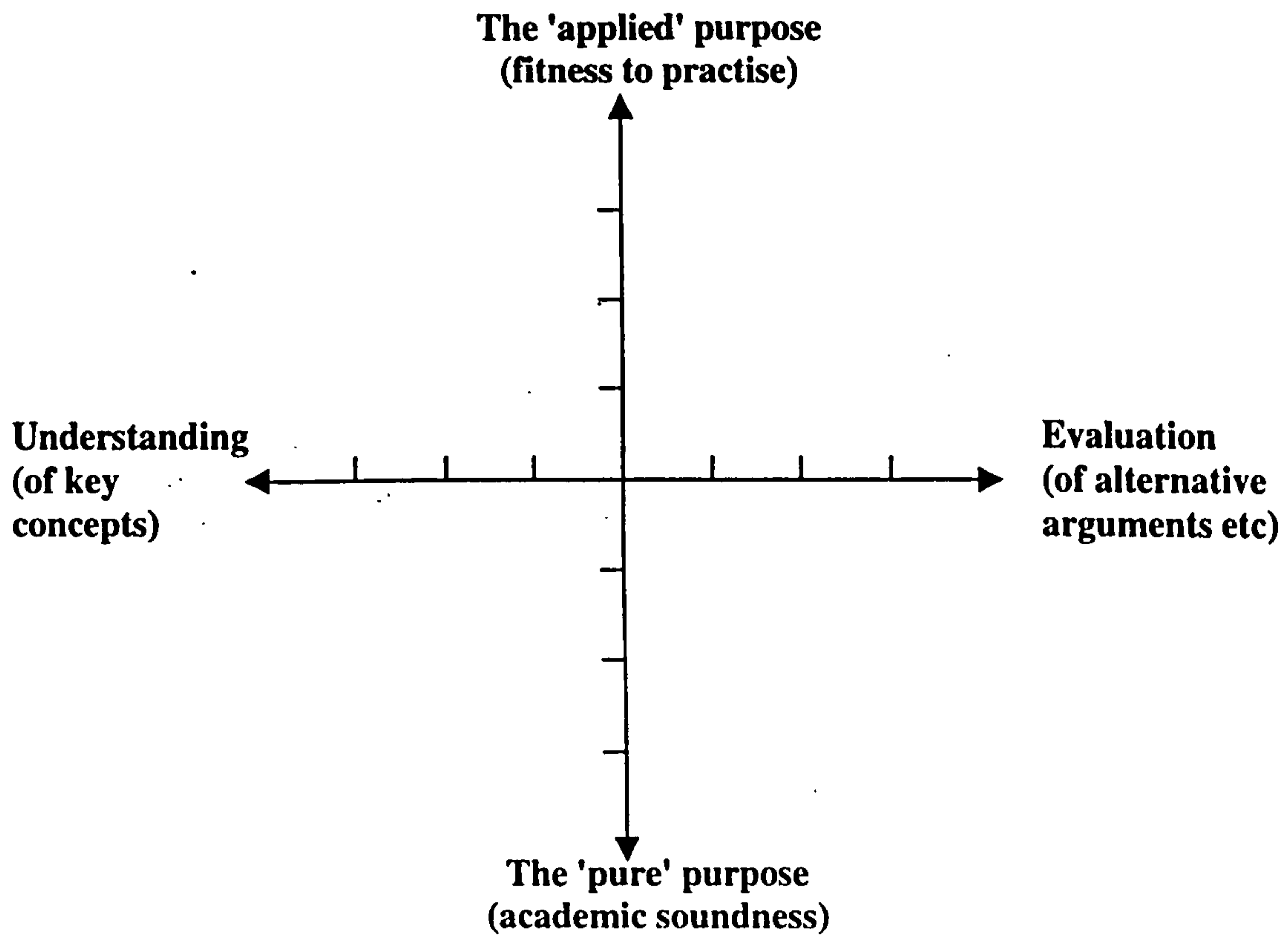
Where does the major pressure on your teaching/assessment time come from?
Teaching commitments, research commitments or what?

H. Is there anything else to do with how design students are currently assessed that you would like to tell me about?

I am most grateful. Thanks very much.

THE CENTRAL PURPOSES OF ASSESSMENT: FINAL (HONOURS) YEAR

Subject area:



	Very important for us	Of some importance	Not important for us
General feeling that change was necessary	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
One 'enthusiast' persuaded us	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Students asked for change (eg S/S comm)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
External examiner pressed for change	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
External validating/professional body	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Increased number of students	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Increased diversity within student population	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Change in course structure	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Faculty policy required change	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Institutional policy required change	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
National policy required change	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Concern for quality of student learning	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Awareness of assessment trends elsewhere	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Educational/assessment 'theorist'	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Awareness of appropriate assessment technology	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>