

Thesis
3928.
(vol. 1)

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

COLIN HOLROYD

VOL I

**A thesis submitted in part-fulfilment of the requirements
for the degree of Doctor of Education.**

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

MARCH 2003

~~06/03~~

DEDICATION

This thesis is dedicated to the memory of Dr Craig Gray of the Teaching and Learning Service of the University of Glasgow. Craig was officially my research mentee but, as suggested in Acknowledgement 2, he was also my mentor. He died in the week of the viva-voce examination of this thesis. He was thirty years old.

**Colin Holroyd
March 2003**

DECLARATION

This thesis has been composed in its entirety by me. The work it embodies has been done solely by me and has not been reported in any other thesis. Parts of the research resulted in internal reports for the University of Glasgow, in a published paper and in a conference presentation. Details of these are given in the thesis and its appendices.

**Colin Holroyd
March 2003**

ACKNOWLEDGEMENTS

I should like the following people to be very sure of my gratitude. All in the first four categories are based in the University of Glasgow.

1. Dr Ron Emanuel, Vice-Principal (Learning and Teaching), and Dr Bob Matthew, Director of the Teaching and Learning Service, approved the research and seemed to welcome its products.
2. All my colleagues in the Teaching and Learning Service put up cheerfully with my cyclothymic behaviour. In particular, Dr Craig Gray (officially my mentee) taught me more than I taught him.
3. The members and convener of the Assessment Working Group forgave me for contributing less to the group than I wished, because of my research role.
4. Ten members of staff were helpful in either preliminary discussions or in pilot interviews. Thirty-eight members of staff agreed to be interviewed and gave me more time than I asked for and more data than I could analyse.
5. Tutors and fellow-students on the Stirling EdD course gave me stimulus and friendship (not necessarily respectively).
6. The love and skills of Mrs Muriel Holroyd ensured that EdD did not stand for *Elegantiae Domesticae Deturbator*.
7. Professor Sally Brown of the University of Stirling was an exemplary supervisor in every way; I respect her too much to embarrass her with all the praise that is her due.

Colin Holroyd

CHANGING ASSESSMENT IN HIGHER EDUCATION: POLICY, PRACTICE AND PROFESSIONALISM

ABSTRACT

The research reported in this thesis focused on the assessment of student learning in higher education. The study aimed to provide practitioners and policy-makers with a research contribution which would increase understanding of change in student assessment by refining simple assumptions about the relationships among policy, practice and professionalism.

The research was carried out in one Scottish university. One strand involved participant observation of a formal group, which had a remit to generate new assessment policy, and documentary analysis of its policy products. The second and major strand was based on semi-structured interviews with thirty-six assessment practitioners in four subject areas (chemistry, philosophy, medicine and design) chosen with the aid of a theoretical model to be as different as possible. These interviews sought the practitioner perspective on significant past changes in assessment (and the reasons for them) and on future changes desired or thought likely to be required. The data were analysed to provide answers to research questions, to identify emerging issues of concern to the participants and to explore imported issues which allowed inferences to be made about the conceptualisation of 'assessment-professionalism'.

The policy group intended to deliver policy-products which would result in greater consistency of assessment practice across the University and generally to

enhance assessment practice. It achieved three main things: agreement on a set of underpinning principles sufficiently broad to allow widely differing interpretations in different faculties/departments; the adoption of a Code of Practice dealing with administrative aspects of assessment and designed to make unacceptable practice less likely; the promotion of policy activity relating to assessment.

Significant past changes were of four types, each associated with a different pattern of causal factors. The types were: evolutionary trends, policy-related shifts, in-course innovations and new-course introductions. The overall amount of assessment change was less than predicted from recent assessment literature. Local innovations within existing courses were very rare. The most striking assessment changes had occurred where new courses had been introduced. Practitioners did not identify policy as a major, direct factor bringing about past changes (except in policy-related shifts), but they expected policy to become more pervasive and prescriptive in future. Policy had a greater **indirect** influence in that it had sensitised staff to the priorities embedded within evolutionary trends and had required assessment to be considered as an integral part of course planning procedures.

Emerging issues showed clear disciplinary differences, but there were common themes in most subject areas. Firstly, epistemological alignment (of assessment with the perceived nature of the subject) was more dominant than constructive alignment (of assessment with educational aims and methods). Secondly, staff were increasingly concerned about the integrity of their assessment methods. Thirdly, the burden of the assessment workload and its management were becoming severe worries. What interviewees said on the imported issues permits the following claims about their assessment professionalism. (a) Assessment was readily accepted as an implicit contractual obligation. (b) The high seriousness of

assessment was acknowledged, but not translated into sufficient time being made available for it. (c) Not all assessors possessed a desirable level of assessment expertise. (d) Assessment practice was not the subject of much critical reflection or creative thought. (e) Commitment to individual ethical action was high, but there was less commitment to communication and interactive professionalism.

The research had some positive impact on both assessment policy-activity and on assessment practice. It contributes to our understanding of how academic staff can be encouraged to engage with important ideas and the links, 'real' or imagined, between them. Participation in the research affected assessment in ways that policy did not.

Future debate on assessment could helpfully centre on (i) the nature of, and effective responses to, student dishonesty in assessment, (ii) encouraging the frequency and depth of communication about assessment, (iii) introducing sustainability into assessment and (iv) the regeneration of academic professionalism around the concept of the academic as educator **and assessor**. There is huge scope for further research in the area; it should include critical policy research, observational studies of professionalism-in-action and attention to the student perspective.

The simple theoretical framework with which the research began was not abandoned, but was elaborated to emphasise that the causes of human action are not single and direct but multiple and interactive.

CONTENTS

	Page
Chapter 1	The Study: Origins, Purposes, Scope and Location
	Origins 1
	Purposes 2
	Scope and research sites 4
	Location: the institution 9
	Thesis structure 10
Chapter 2	Conceptual, theoretical and contextual aspects
	Introduction: theory guidelines 12
	Primary concept 1: policy 15
	Primary concept 2: assessment practice 21
	Primary concept 3: academic professionalism 25
	Secondary concept 1: change and response to change 35
	Secondary concept 2: disciplinary difference 38
	Relationships between concepts 39
	The national policy context 44
Chapter 3	The Research Activity
	Aspects of planning (1): the participants 48
	Theory: an interpolation 49
	Aspects of planning (2): research approach and methods 52
	What research was carried out? When? 54

Chapter 4 Changing Assessment Policy

The Assessment Working Group: activities and reports	65
The nature of the assessment reports	69
The Assessment Working Group: critical comment	76
The perspectives of two key informants	81
Did the University generate ‘new’ assessment policy?	88
Did policy aim to change assessment practice?	89
Coda	90

Chapter 5 Changing Assessment Practice: Chemistry

Introductory note	92
Was chemistry seen as one subject area?	92
How did participants locate chemistry within the ‘four quadrants’ framework?	92
What significant changes had occurred? Why?	93
What further changes (a) did staff wish to see and (b) think might be required of them?	100
Was policy perceived as influential in bringing about change?	103
Emerging issues	104
Imported issues	107
Evaluative comment from interviewees	111

Chapter 6 Changing Assessment Practice: Philosophy

Was philosophy seen as one subject area?	113
How did participants locate philosophy within the 'four quadrants' framework?	113
What significant changes had occurred? Why?	114
What further changes (a) did staff wish to see and (b) think might be required of them?	120
Was policy perceived as influential in bringing about change?	123
Emerging issues	124
Imported issues	128
Evaluative comment from interviewees	131

Chapter 7 Changing Assessment Practice: Medicine

Preamble	133
Was medicine seen as one subject area?	133
How did participants locate medicine within the 'four quadrants' framework?	133
What significant changes had occurred? Why?	134
What further changes (a) did staff wish to see and (b) think might be required of them?	142
Was policy perceived as influential in bringing about change?	149
Emerging issues	150
Imported issues	155
Evaluative comment from interviewees	159

Chapter 8	Changing Assessment Practice: Design	
	Preamble	160
	Was design seen as one subject area?	162
	How did participants locate design within the 'four quadrants' framework?	162
	What significant changes had occurred? Why?	163
	What further changes (a) did staff wish to see and (b) think might be required of them?	168
	Emerging issues	171
	Imported issues	175
	Evaluative comment from interviewees	178
Chapter 9	Answering questions and discussing issues	
	Introduction	180
	Section 1: Locating Subjects in the four quadrants	180
	Section 2: Changing assessment practice	184
	Section 3: Relating assessment policy and practice	203
	Section 4: Issues emerging and imported	212
	Section 5: The impact of policy and practice on professionalism	231
	Section 6: Personal reflection on salient issues	243

Chapter 10: Revisiting the theoretical framework and reflecting on the research

Revisiting the theoretical framework 248

Reflecting on the research 258

Was the research well-done? 259

Was the research worth doing? 266

What should happen next? 274

REFERENCES 276

APPENDICES Separate volume

CHAPTER 1

THE STUDY: ORIGINS, PURPOSES, SCOPE AND LOCATION

ORIGINS

The Doctor of Education of the University of Stirling is a research degree with a taught component. It is officially described as having a more professional orientation than the traditional Doctor of Philosophy. In the formal documentation, the aims of the doctoral course are to develop understanding (a) of what research can offer practitioners and policy-makers in their decision-making, (b) of the power that research has to explain why things are as they are and (c) of the policy implications of research. Although methodological and substantive concerns run in parallel throughout the four modules of the taught component, the titles of two modules indicate a methodological emphasis (*Analysis, criticism, interpretation and use of research, Research planning and design*) and of the other two a substantive emphasis (*Institutional change as a field of study, The impact of policy on professional practice*). When the present writer encountered these latter modules teaching was organised around three distinct foci: understanding institutional change, the impact of policy on practice and the nature of professionalism.

The writer has had a strong personal interest in the assessment of students in higher education over a period of forty-seven years as a person assessed, as an assessor of others and as an assessment developer. Although he could probably produce evidence of being judged adequate in these three roles, he has become increasingly convinced that, given its crucial role in higher education, assessment is still not well-enough understood, certainly by him, but also by those who have consulted him in recent years about how assessment might be improved. Also,

assessment (in higher education) is still a relatively under-researched area. This writer is not alone in holding these views. For example, Wakeford (1999) states that 'The assessment of students' learning is an under-discussed and, in most disciplines, an under-researched aspect of higher education'. Yorke (2001) concludes that 'assessment is, by general consent, the least well-understood and enacted aspect of [higher education] curricula'.

When it came to designing the research reported in this thesis, it was decided to embrace explicitly the aims of the Stirling EdD programme, the substantive foci of the taught component and the personal interest in the assessment of students in higher education.

PURPOSES

The research had three broad aims:

- (i) to increase understanding of why, in the assessment of students in higher education, things are as they are and why things change,
- (ii) to explore relationships among assessment policy and practice and professionalism and
- (iii) to offer assessment practitioners and policy-makers a research contribution with implications for practice and policy.

To guide the research and to make it manageable in scope, research questions were formulated. The initial attempts were tentative and provisional and went through a process of evolutionary change; this can be traced in assignments submitted in 1999 and 2000. (Appendices I and II)

1. Did the University of ¹Glasgow generate new assessment policy between 1998 and 2002?
2. Did policy aim to change assessment practice?
3. What did academic staff in different subject areas think was good about how students were assessed?
4. What significant changes had taken place? Why, in the view of staff, had these changes occurred?
5. What further changes did staff (a) wish to see and (b) think might be required of them? Why?
6. To what extent was policy perceived as being influential in bringing about assessment change?
7. What impact did changing assessment policy and practice have on academic professionalism?

Questions similar to these, with appropriate differences of tense, were all formulated before the research activity began. As the work progressed, a further research question emerged. (Page 8)

8. Did members of staff locate their own subject areas as predicted from the theoretical model which guided the choice of subject areas for the research?

Some would recommend that research purposes be expressed as hypotheses rather than questions. The present writer prefers not to do this and the reason should be made clear. If an hypothesis is taken to be any general statement, either

¹ This naming of the institution in which the research was conducted is in line with the 'rules of engagement' devised for the research. (Chapter 3)

descriptive or explanatory, that can be formulated in advance of the research, then it follows that the appropriate thing for the research to do is to test that hypothesis. Used in this way the word hypothesis carries necessary implications of prior formulation and testability; it thus adumbrates a research approach that will be essentially positivist and concerned with theory testing.

Research question 8 above can quite reasonably be re-phrased as an hypothesis: 'Practitioners will locate their subject areas as predicted by the framework'. This is a testable hypothesis and indeed it is one that was tested during the research. On the other hand research question 4, 'Why, in the view of staff, had these assessment changes occurred?' cannot appropriately be re-worded as an hypothesis. It would be illegitimate for the researcher to formulate in advance of the research how the question would be answered. There is as yet no hypothesis to test and answering the question requires other than a positivist stance from the researcher.

This reluctance to provide formal statements of hypotheses at the outset does not deny the existence of 'weak' hypotheses (or hunches) about what might emerge from the research. The idea that any researcher can embark on research equipped with nothing but questions is naïve.

The position taken above, that some of the research appears positivist but most is not, raises questions about consistency in the theoretical perspective underpinning the research approach. These are addressed in Chapter 3.

SCOPE AND RESEARCH SITES

In Scotland, there is little tradition of research into student assessment in higher education. The only researchers based in Scotland and known to this writer

were Hounsell and Falchikov in Edinburgh (e.g. Hounsell and McCulloch, 1999; Boud and Falchikov, 1989), Stefani in Strathclyde (e.g. Stefani, 1998) and Mason in St Andrews (Mason, 1998, 2000). A literature research revealed no other significant sources.

There has been no Scottish research, with a theoretical basis, that has been national in scope. This may seem unfair to the work of Hounsell and his colleagues in producing the ASSHE Inventory of Changing Assessment Practices in Scottish Higher Education. (ASSHE, 1996) This was certainly national in scope and the inventory remains a very useful document; however, it was the outcome of assiduous information gathering which did not fulfil the original research intentions. In describing the aims of the ASSHE project, the authors state that it is 'specially important not only to identify the myriad ways in which [assessment] practices are actually changing, across subjects and institutions, but also to pinpoint factors which seem to work for or against particular changes'. This was the original intention but, in a personal communication to the author, Hounsell described his regret that his project did not in fact have the time or resources to move beyond identifying and describing assessment changes.

This writer decided reluctantly that to attempt research on a national scale as a sole researcher would mean sacrificing depth for breadth. Serious consideration was given to carrying out case studies in two contrasting institutions, the Universities of Glasgow and Stirling. This idea had obvious methodological attractions, but it was abandoned (again reluctantly) when it was realised how severely the research focus would have to be restricted for the empirical work to remain manageable.

At this point in the planning, the writer was unexpectedly invited to become an Honorary Senior Research Fellow within the Teaching and Learning Service of Glasgow University. He was pleased to accept and decided that the proposed research should be sited within the institution providing him with a base. It should be emphasised that the honorary position meant the provision of a desk; it did not mean receiving any research funding or any secretarial support. Nor was there any stipulation that the research should be useful to, or indeed comfortable for, the institution.

The aims of the research were discussed with, and approved by, the Vice-Principal (Learning and Teaching) and the Director of the Teaching and Learning Service. Thereafter there was no intervention in the conduct of the research or in the content of any report. Apart from the informal support of immediate colleagues, the researcher was on his own, with all the advantages and disadvantages that implies.

Thought was then given to the feasibility of taking Glasgow University as a whole as one site for a case study. After consultation with others, it was decided that this was unnecessarily ambitious for the purposes of this thesis, but the idea was not dismissed. Proposals were developed for Plan A to cover the period until 2002 (research for the Stirling EdD) and Plan B (which would include all Plan A activities) to cover the period until 2004. (See Appendix I)

Plan A The minor (policy) strand – participant observation of a policy-generating body and interviews with key policy-makers.

 The major (practice) strand – an interview-based study in four subject areas.

Plan B The work of plan A would inform an institution-wide survey by questionnaire.

The survey would allow the identification for interview of further key people in changing practice and policy.

The research 'sites' for the minor strand were very simply decided. The author had been invited to join the Assessment Working Group, a committee with the remit of generating assessment policy. The member of senior management responsible for setting up the group and the convener of the group were obvious candidates for interview.

It was difficult to decide the sites for the major (assessment practice) strand. It seemed likely that a minimum of about eight interviews would be necessary to gain understanding in adequate depth of the range of viewpoints in each subject area and that about thirty interviews would be practicable. Four different subject area sites would be reasonable, but which four? There was already an assumption that there would be subject-based differences in assessment practice and thinking, so it seemed desirable that the subject sites be as different as possible.

Current writing on disciplinary differences in assessment tends to distinguish subjects by their position on a hard-soft dimension. Examples of this are to be found in Mason (1998), Bridges et al (1999) and Yorke et al (2000). Mathematics and computing are located at the 'hard' end of a spectrum, biology and social sciences somewhere in the middle and fine art and literature at the 'soft' end. This conceptualisation was judged inadequate because it left out too much, especially the differences between those subjects which were essentially 'pure' or academic in intention and those which were 'applied' or broadly vocational. It was

thus posited as a working hypothesis that subject areas could be positioned with respect to two orthogonal axes, the hard-soft and the pure-applied. This framework had the obvious practical advantage that it pointed to the desirability of selecting four subject areas, one from each quadrant of the framework. There was an immediate research consequence; the framework itself should be tested within the research. Would participants accept the framework as valid and would they locate their subjects in the quadrants predicted by the researcher? (Research question 8)

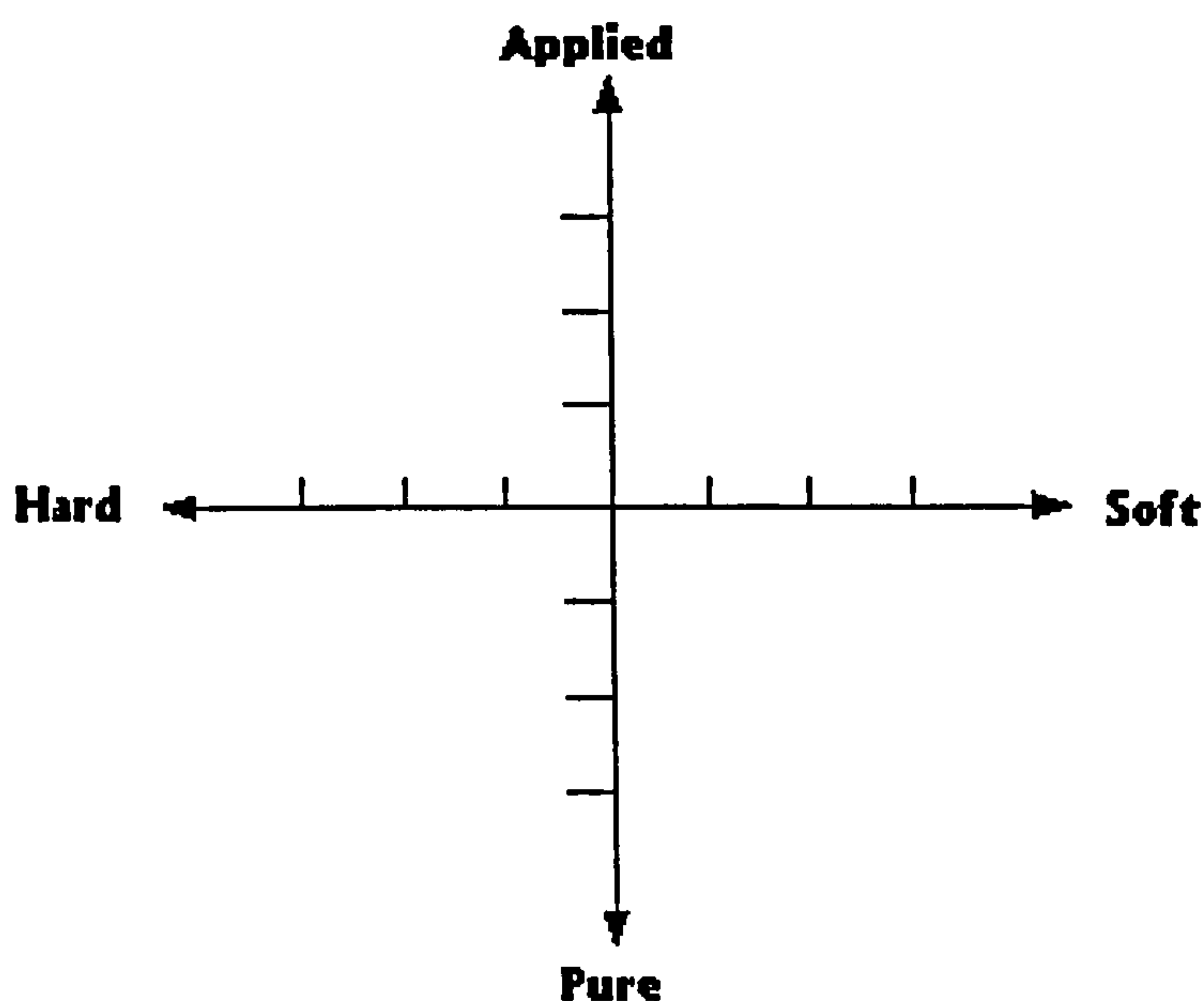


Fig. 1 The 'four quadrants' framework

Guided by this framework, four subject areas were chosen: chemistry (hard and pure), philosophy (soft and pure), medicine (hard and applied) and design (soft and applied). Clearly the framework did not prescribe these subjects in preference to say, physics, literature, engineering and law; other influences were at work. The writer was originally a chemist and thought continuing contacts would simplify access negotiations; he had worked with the Medical Education Unit and knew that assessment procedures had changed radically; he had been keenly interested in

philosophy and in design but knew little of their assessment procedures and was keen to learn more. The choice of design as a subject area meant a welcome opportunity to work in Glasgow School of Art, an associated college of Glasgow University

LOCATION: THE INSTITUTIONAL CONTEXT

In 1451, King James II of Scotland persuaded Pope Nicholas V to grant a Bull authorising the setting-up of a university in Glasgow. Thus, 40 years after the creation of St Andrew's University Scotland, like England, could boast two universities. Modelled on the University of Bologna, Glasgow was (and claims to have remained) a university in the great European tradition. Today Glasgow University is a large and complex institution. In session 2001-2002, it had 15,200 undergraduate and 4,000 postgraduate students; there were 5,900 members of staff of whom 1,600 were academic staff. There were eleven faculties embracing about 100 departments and 70 centres or schools. Although there has been a Department of Education for many years, there has been no tradition of research into higher education. The Teaching and Learning Service was established in 1994 as a central service without a research function; it became part of a new Faculty of Education in 1999; a Centre for Research in Higher Education was set up within the Teaching and Learning Service in 2001.

Glasgow School of Art was founded in 1845. At the heart of the present campus is Charles Rennie Mackintosh's masterwork, now known as the Mackintosh Building, a place of pilgrimage for visitors from all over the world. The school is an associated college of the University of Glasgow and its degrees are accredited by the University, but the School of Art remains in some ways a distinctive institution. In session 2001-2002, it had a total of 1,400 students and 100 full-time academic

staff. There were three component schools (fine art, design and craft, and architecture) and one general underpinning department, historical and critical studies.

THESIS STRUCTURE

The conceptual and theoretical aspects of the research are dealt with in Chapter 2 (and Appendix III) and then re-visited in Chapter 10. References to relevant substantive literature are concentrated in these two chapters and in a theory interpolation in Chapter 3. Methodological references for the research approach and methods occur mostly in Chapter 3, an account of the planning and conduct of the research.

The minor (policy) strand focused on research questions 1 and 2 and is reported in Chapter 4. The major (practice) strand addressed research questions 3, 4, 5, 6 and 8. Chapters 5, 6, 7 and 8 deal with changing assessment practice in the four subject areas (chemistry, philosophy, medicine and design respectively) and address questions 4, 5, 6 and 8. This appears to leave research question 3 unanswered. This question (about what practitioners found pleasing in their current practice) was essential to the conduct of the research; however, given the title of the thesis (which limits the focus to **changing** assessment practice) and the restriction in word length, it was decided not to answer the question in the main body of the thesis. Full answers have been relegated to Appendices IV, V, VI, VII and VIII – the reports provided to the University of Glasgow.

Chapter 9 is a lengthy chapter. Firstly, it pulls together participants' views on the location of their subjects within the 'four quadrants' framework (Research Question 8). Secondly, it provides and discusses summary general answers to

research questions about changing practice (RQs 4 and 5). Thirdly, it pulls the policy and practice strands together (RQs 1, 2 and 6). Fourthly, it looks at the issues emerging from, and imported into, the major activity of interviewing. This discussion leads, fifthly, into a consideration of the relationships between policy and practice, policy and professionalism and professionalism and practice (RQ 7). The chapter ends with a personal reflection on four salient issues.

Chapter 10 re-visits the theoretical framework and re-examines the 'early notions' of conceptual relationships introduced in Chapter 2. Critical reflections on the design and conduct of the research and some consideration of its impact conclude the thesis.

CHAPTER 2 CONCEPTUAL, THEORETICAL AND CONTEXTUAL ASPECTS

INTRODUCTION: THEORY GUIDELINES

The writer devised six 'theory guidelines' for the design, conduct and writing-up of the research. Each guideline was derived from answers to questions that he felt appropriate for this research. Brief answers are given here; Appendix III provides an extended treatment.

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

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

The research aims to explain 'why things are as they are'. Description is taken as prior to, and necessary but insufficient for, explanation. Evaluation and prescription are **not** purposes of this research.

(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?)

Concepts are taken to be general abstract ideas. A set of concepts forms the basis, organiser and provider of significance. The metaphor of 'basis' is necessary, but can mislead; the metaphor of 'superstructure' is a helpful supplement.

(c) *Theory is taken into the 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 look like in this context? How are the theoretical and conceptual frameworks related?)

In this research, a theory is taken to be a statement which expresses an ordered relationship between concepts and which has some explanatory value. The theoretical framework is seen as a network in which concepts form the knots and relationships between concepts are the threads. A fuzzy, incomplete framework is taken into the research; one aim of the research is to make a modest but creative contribution to clarifying the nature of both threads and knots.

(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 theoretical framework of the content?)

The research approach is ethnographic. The research will, however, attempt to make explicit the theoretical notions taken into it and this may debar it from ethnography in its purest, most restricted form. (See Appendix III, page 9.) The methods used within this approach are semi-structured interviewing, participant observation and thematic documentary analysis. The theoretical perspective underpinning the approach is interpretivist or, more precisely, symbolic interactionist. Some qualification of this last term may be necessary. There is a tendency for symbolic interactionism to slide from its central concern with the cultural provision of meaning into cultural determination. This research will proceed on the assumption that individuals retain potency and agency as actors within a culturally-provided social drama. (See Appendix III, page 8.) The underlying epistemology is constructionist. This description is based on terms as defined by Crotty (1998).

Any particular area can be studied by a whole range of approaches; the theoretical perspective of the approach does not determine the theoretical framework of the substance. In this research, the theoretical perspective of the methodology does not influence the choice of concepts, but it does influence the nature of the knowledge claims that are made about the linkages between concepts.

(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 research will distinguish between two kinds of issue. Imported issues (or etic issues – see Stake, 1995) are ones that the researcher has in mind before the research and wishes to explore during it; emerging (or emic) issues are unpredicted ones that emerge from the research as important to the participants.

(f) The research must be located in context.

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

There are four important aspects of context: the time, the place, the policy environment and the theory framework. The time context is dealt with as and when appropriate throughout the thesis; the place (institutional context) was described in Chapter 1; the theoretical and the national policy aspects are provided in this Chapter.

Readers are reminded at this point of the thesis title: ‘Changing assessment in higher education: policy, practice and professionalism’. Title and theory guidelines influence the structure of the rest of this chapter.

PRIMARY CONCEPT 1: POLICY

Dictionary definitions are easy to provide. For example, policy is ‘a course or principle of action adopted or proposed by a government, party, business or individual’. This doesn’t take us very far.

In an exceptionally helpful book on the concept of policy, Colebatch (1998) notes that this problem of definition has been a source of concern to many influential writers in the field. Some have felt that definitional questions just get in the way and have given up endless ‘definitional jostling’ to get on with substantive policy research. This present writer would argue that the premature imposition of an exclusive definition might well prevent researchers seeing important meanings that the term ‘policy’ can carry. Equally, the attempt to avoid **any** conceptual clarification aiming at some preferred, albeit temporary, definition would be an evasion of research responsibility (unless the sole purpose of the research was to explore the definitions used by others). Ambiguity about definitions may at times be helpful in policy activity, but not in policy research.

Colebatch’s explanation of the difficulty in reaching any satisfactory definition of policy is that, both for academic researchers and practitioners, there are two fundamentally different perspectives on policy that frame policy action in divergent ways. The first sees policy as **authorised choice**; this assumes that policy is about ‘governments making decisions’ and it focuses on those decisions. This perspective asks questions about the problem the government was trying to address, the options it considered, how it made the decisions it did and what the outcome was. Colebatch first describes this perspective as applying to government, but then extends it to governing bodies in general, for example to the legitimate ‘authorities’ within higher education institutions. The second perspective sees policy as **structured**

interaction. This does not assume a single decision-maker addressing a clear policy problem. It focuses instead on the range of participants in the policy 'game', the diversity of their understandings of the situation and the problems framed from it, the ways they interact with each other and the outcomes of their interaction. It does not assume that this pattern of activity is a collective activity in pursuit of known and shared goals.

There exists a rational model of policy-making which sees it as a cycle of applied problem-solving: setting the agenda, policy formulation, decision-making, policy implementation and policy evaluation. This model is not a good description of what actually happens, but it is significant simply because it is in use. The challenge thus is to come up with a definition which is satisfactory in that it accommodates the tension between the model and how it is used. Colebatch offers his own definition: 'Policy is a term used to refer to the structuring of collective action by the mobilisation of a model of government as authorised decision-making'. This is an awkward definition, but it does draw attention to three important elements.

Policy is process as well as artefact. In common usage policy is an artefact, a thing created by policy-makers; it does not exist unless embodied in an official document. But formal policy production can only be understood in terms of that process which embraces all the action which takes place around the possibility of some use of authority to structure action; policy statements and artefacts like ministerial speeches, White Papers, and institutional policy documents are all part of the process of structuring. Some would counter this with the argument that clear policy products do exist and they need to be distinguished from all the ordinary

processes of government and management. In addition, policies have objectives and it is reasonable to ask whether or not they are achieved.

This latter argument is valid but incomplete. Policy products have to be understood in the context of other stated policies and of broader factors that also influence and structure action. Whether a policy is judged effective depends on how this question is framed and by whom, and who is asked to answer; this too is about process.

A concern for the process aspects of policy then raises the question of whether other forms of structuring action, ones which do not involve authorised decision-making, should also be seen as policy. For our present purposes, it is useful to return to Colebatch; he discusses this issue through an example concerned with professionalism. Teachers' actions in deciding which children in their care should be punished, when and how, have in the past been governed solely by the exercise of professional judgement; when teacher action comes to be governed by an explicitly stated 'discipline policy', there has been a highly significant change in the structuring of action. The structuring of actions by 'professionalism' does not amount to 'policy' because of the absence of the required feature of authorised choice.

There is a potential complication here. Could a professional body decide the authorised choices? Could professionalism be seen as policy? Professions have traditionally stood outside the chain by which the legitimate authority of government is delegated and exercised; they have argued that there are essential aspects of their work where there must be scope for discretionary judgements that they can be trusted to make. To be told what to do would then imply the loss of a cherished independence and autonomy. However, the actions of professionals have always been constrained in some aspects by law and by policy – whatever the source of these. The extent to

which these constraints are seen as welcome or even acceptable does seem to depend on the perceived source of policy: policy from the government can be seen as ‘unwarranted interference’ when the same policy from the professional body would be seen as ‘reasonable self-regulation’.

Policy is concerned with creating coherence in the face of continuing ambiguity and contest. There may be a shared picture of decision-making as a clear process, but the experience is one of contest, ambiguity and confusion. This can at least partly be understood by conceptualising policy as operating on two dimensions, the vertical and the horizontal. In the former ‘policy’ is construed as the transmission downwards of decisions from the top. Authorised decision-makers choose courses of action which maximise the values they hold and transmit these to subordinates to implement. This dimension stresses instrumental action, rational choice and the force of legitimate authority. There can be upward movement, for example of the results of evaluation of policy implementation, which may or may not inform future policy-revision. On the other hand, the horizontal dimension is concerned with relationships and interactions amongst policy participants in different organisations, some outwith any overt line of hierarchical authority. Policy work takes place across organisational boundaries as well as within them; it is concerned with linkages across units of organisation, how they are formed and maintained, with the interpretive frameworks with which participants understand the policy issues and the institutional strategies within which these are mobilised.

The two dimensions are not alternatives; rather, each assumes the other. They offer very different answers to the question ‘Where is policy made?’ In the vertical dimension, policy is made when the authorised decision-makers give their assent in

ministerial offices, the Cabinet Room, or parliament (or within a university, within the principal's office, the senior management group, the Court or the Senate). In the horizontal dimension, policy emerges from a complex set of relationships amongst participants, marked by continuity and ambiguity rather than clear choices; policy is not so much 'made' as continuously formed and then re-formed.

The horizontal dimension of policy is a field always marked by tension, ambiguity and contest. (These aspects are highlighted by Ozga (2000) in her book, *Policy Research in Educational Settings*. The sub-title is *Contested Terrain*.) There will be tension between the vertical and the horizontal dimensions, for instance between articulating clear purposes and procedures and accommodating the priorities of different participants. There will be ambiguity arising from different perspectives, from the (perhaps deliberate) imprecision in language used and from the gap between the preferred discourse of participants and that of decision-makers.

Policy is always problematic and graduated rather than definitive and absolute. This follows from the previous two assertions; it is an inevitable consequence of construing policy (a) as both process and artefacts, (b) as both the contested terrain of policy-activity and the linear model of policy-making, and (c) as a socially constructed variable rather than a 'scientific' absolute. There is a temptation to argue that policy will be particularly problematic and graduated where there is a tradition of autonomy (and intelligent awkwardness) and where currently there is real uncertainty as to what form of professionalism ought to be claimed i.e. within our higher education institutions.

This discussion of the concept of policy has emphasised that policy is not only national (or supra-national) public policy (deriving from government). For example, a university is not a simple unitary body on the receiving end of policy directives from one clear government source (although these clearly exist and are increasing); rather it is subject to a constellation of policy influences and requirements, some of which require it to generate new policies of its own, which then add to the policy influences on the component parts of the university. Bauman (1999) helps illuminate this.

Where is policy? It is in the ecclesia, the agora and the oikes. The ecclesia are the official sites of policy determination (government, assembly, senate). The agora are the meeting places in which ideas are traded, 'the arenas of agonistic drama'; in higher education these are both the formal inter-university and intra-university committees and working groups and related conferences and seminars. The oikes ('homes') are the more private and domestic sites. Within a university these are departments, course teams, boards of examiners and staff/student encounters. Some individuals travel between ecclesia and agora, and between agora and oikes; but not all do. Policy officially comes from ecclesia; but the more detailed policy and the strategies for its implementation develop within the agora in ways more or less consistent with the original policy outline. It can then be predicted that the people (academic staff) in the oikes will be uncertain as to the 'real' source of policy. Was 'the university' the source or the channel? Bauman also helps our understanding of what happens to policy in the oikes; these are not merely reception sites, in which the policy is applied or implemented, they are sites of redefinition, reconstruction and resistance. Policy exists for the structuring of actions, i.e. to have an impact on practice, but practitioners can influence policy, not only in the provision of feedback about the effectiveness of implementation, but in their constructions of the meaning

of the policy. Practitioners may rarely be the policy-makers, but they are always players in the drama of policy activity. To emphasise that policy is not merely a product received, but a complex drama in which individual actors engage, is to echo the stance adopted by Ball. Whatever policy is, it is 'not passively received and automatically implemented, rather it is actively interpreted, decoded and responded to in complex social and cultural contexts'. (Ball, 1994)

In the previous paragraph the impact of policy on practice was explicitly mentioned. It should be noted that any consideration of the impact of policy on practice cannot altogether avoid considerations of power. The writer made a reluctant decision not to pursue this facet of the research.

PRIMARY CONCEPT 2: ASSESSMENT PRACTICE

What is assessment? The question seems an obvious one to ask, but many authors of influential texts on the assessment of students in higher education do not attempt to answer it. For example, Brown et al (1997) provide an etymological justification for what they believe assessment **should** emphasise. The term comes from the Latin 'ad sedere' which means to sit down beside someone and therefore 'assessment should be primarily concerned with providing guidance and feedback to the learner'. Brown and Glasner (1999) start from the assertion that assessment should be an integral part of student learning. The meaning to be attached to the term is also assumed to be non-problematic by Hounsell and colleagues (ASSHE, 1996); they go straight into a description of changes in assessment practice. In his book 'Assessing Student Centred Courses', Gibbs (1995) finds it necessary to present and discuss at length definitions of student-centredness, but not of assessment itself. The twelve

booklets making up the Assessment Series tackle a huge range of questions about assessment, but do not address the matter of definition. (LTSN, 2001)

Intriguingly, texts on assessment not focused on higher education generally do address the problem of definition. Gipps (1994) defines assessment as 'a wide range of methods for evaluating pupil performance and attainment'; for Harlen (1994) assessment is 'the process of making judgements about a student's performance in particular tasks'; Broadfoot (1996) defines assessment as 'the deliberate and overt measurement of educational performance to provide information for purposes beyond the immediate interactive learning situation'.

Of course some texts on assessment in higher education do define the term. Consider just three examples, one classic and two recent. Rowntree (1977) defines assessment as 'getting to know our students and the quality of their learning'. Freeman and Lewis (1998) take the main current meaning of assessment as being 'an estimation, judgement or evaluation of the extent and worth of student learning'. Heywood (2000) defines assessment as 'a multi-dimensional process for judging individuals in action'.

Wolf (1995) writes in her first paragraph that 'rather than drag the reader on a tour of these definitions, their advantages and failings, I want to start with a definition of my own'. The present writer will follow this example. For this thesis, assessment will be defined as **any process which allows the making of a judgement about the extent and nature of student learning**. Four comments on this are needed.

- (i) In a policy document to be described later (Chapter 4), a working group defined assessment in very similar terms to this but prefaced 'judgement' with the

adjectives 'professional, academic'. The extent to which assessment judgements should be seen as academic and professional is a theme in Chapters 9 and 10.

(ii) Assessment can be both the administration of an assessment and the making of a judgement about student learning as a result of that process. The former is often quite easy, but the latter is intrinsically and unavoidably difficult. The reason for this should be clear: the judgement is always an inference about the internal mental state ('learning') of a complex human being from the observation of some sample of behaviour or of an artefact (performance or product). Although in some assessments of 'simple' skills and competences the necessary inferential leap is relatively small, in judging subtle intellectual capabilities it is very large indeed.

(iii) Assessment is a much broader concept than examination; examinations are one form, or mode, of assessment but also included are a huge number of other forms: essays, project reports, theses, laboratory work, clinical performance, oral presentations, portfolios of work and so on. Assessment embraces both formal and informal episodes; students may experience everything from a high-stakes viva-voce examination to the kind of brief, low-stakes informal interaction with a tutor which is indistinguishable from teaching.

(iv) The forms that assessment can take should be distinguished from the purposes they may serve. There are many explications of the latter. See, for example, Atkins et al (1993), Freeman and Lewis (1998 – pp 10-11), Miller et al (1998 – Part 1) and Brown and Glasner (1999 – p 6). Long lists of purposes are sometimes categorised into three broad purposes: certification, accountability and learning enhancement. The first two of these are on occasion conflated into certification/ accountability (assessment for the public or the 'control' function of assessment); learning

enhancement is then described as assessment for the individual (the 'growth' function). (Nisbet, 1993; Broadfoot, 1996)

'Practice' is that set of actions, executions and procedures that a person actually performs and is to be distinguished from theory – what they say and think about what they do. In talking of practice there is a tendency to talk of actions which are learned, repeated, habitual and routine. Of course there are repeated patterns of action in any person's practice, but the use of these adjectives is dangerous if it suggests a division between non-intelligent practice and rational theorising. Although people sometimes think before (and after) acting, 'it is also true that in much of the spontaneous behaviour of skilful practice we reveal a kind of knowing which does not stem from a prior intellectual operation'. (Schön, 1983 – p 51) In Gilbert Ryle's words, 'When I do something intelligently, I am doing one thing not two'. The idea of practice as the reproduction of routines is also misleading. In the realm of medical practice, for example, influential research emphasises that the quality of the practitioner lies not merely in the use of a large repertoire of appropriate internal 'scripts', but also in the ability to recognise those situations in practice for which no script exists. (Schmidt et al, 1990)

With the caveats outlined, it appears relatively easy to move from the concepts of assessment and practice to assessment practice; assessment practice is what assessors actually do. The core activities must be conducting assessments and making judgements about the extent and nature of student learning. There are, however, other elements of practice (pre-active, interactive and post-active) which may have to be included, elements which enable the essential ones or which stem from them. (Yorke, 1998; Knight, 2002) An indicative, but not exhaustive, list includes:

- (a) providing feedback to students on the basis of assessment judgements,
- (b) aggregating individual assessments to a summary judgement about standard of performance in a course or programme,
- (c) collaborating with others e.g. in double marking or in boards of examiners,
- (d) managing an assessment system (to include the preparation of assessment instruments and episodes, the recording and reporting of results, appointment of and liaison with other internal and external examiners),
- (e) innovating in assessment and integrating assessment into the planning of new courses.

PRIMARY CONCEPT 3: ACADEMIC PROFESSIONALISM

Professionalism is a strangely fuzzy, protean concept. At a recent seminar attended by this author an eminent academic provided a penetrating analysis of the concept of traditional professionalism and then provided a persuasive argument in favour of his own preferred re-conceptualisation of the term. A member of his audience erupted: 'This is futile faffing around; the answer is simple – professionals get paid for what they do, others do not'. At another seminar, one speaker said that there were now so many ill-defined meanings of the term professionalism that it had become essentially meaningless: 'Use of the term is unnecessary and unhelpful and should be banned altogether'. As a preliminary to this section, both of these reactions will be scrutinised.

Understanding of the first reaction requires a distinction to be made between 'being a professional' and 'being professional' i.e. behaving in a professional way.

The opposite of being a professional is being an amateur; the former is paid, the latter is not. The opposite of behaving in a professional way is being unprofessional; the former then implies that the behaviour is in some way worthy of commendation, the latter implies disapproval and censure. If the only requirement of professionalism was that people be paid, then all paid occupations would become professions and there are powerful objections to this. (Is football to be regarded as a profession merely on the grounds that professional footballers are paid? Even those who would argue this case, baulk at the idea of book-making and drug-dealing as professions.) Being paid seems to be a necessary but not sufficient condition of being professional. But there is a further complication in that people can be expected to behave professionally even when unpaid. A voluntary social worker is at least a quasi-professional; so too are medical and education students on placement. The essential requirement is not simply payment, but that the person agrees to perform certain tasks and accepts accountability for quality of performance.

The second reaction ('the concept is unhelpful – dismiss it') is understandable. There have been many crises of professionalism; the professions have often failed to live up to public expectations of them; the concept itself is messy, slippery and mutable. However, it does seem to be the case that there is strong and widespread pressure for its retention across education; the concept will not just slink away. Two recent examples illustrate the point. Firstly, Walker (2001) and co-authors were highly critical of that dominant professionalism which sought 'not to rock the boat'; however, they decided they did not wish to dismiss professionalism, but to contest and re-work what it means to be professional in higher education. Secondly, The Institute for Learning and Teaching in Higher Education is 'the professional body for all who teach and support learning in higher education on the United Kingdom'; it

aims 'to develop and monitor professional standards of practice'; it is becoming 'the main source of professional recognition' for all academic staff. (ILT, 2000)

The extensive literature on professionalism variously describes it as traditional, quintessential or residual, as semi-, restricted or extended, as dispersed, hybrid or syncretistic, as corporatist or bureaucratic, as dominant or emergent, as re-imagined, re-theorised, re-defined, re-conceptualised, regenerated, reconstructed, reborn or reclaimed. The following analysis is a summary version, with extension, of what has already been written in Assignment 3 (Appendix IX) and in a paper entitled 'Are Assessors Professional?' (Holroyd, 2000; Appendix X) Four ways of conceptualising professionalism, one traditional and three emergent, will be covered; there is then a suggestion as to how these four could be condensed to serve the purposes of this thesis.

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 some form of higher education which involved the mastery of some difficult body of useful knowledge and some form of extended 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 of the professional and of the client or manager meant that 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.

Schön (1983) encapsulates some of this by writing that 'the traditional function of the professional was to apply specialist knowledge with vigour, probity and community orientation'. Freidson (1994) attempts to summarise all the characteristics of traditional professionalism into two basic requirements: 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 professional status for financial gain; the fragmentation of existing professional groups; growing general public distrust of experts; increasing managerialism and state control; and a re-theorising of specialised knowledge. The period since has witnessed the emergence of new concepts of professionalism which their adherents argue are better adapted for survival in a post-modern environment.

The first of these can be labelled 'communicative' professionalism because of the importance within it of various types of interaction. The important author here is Nixon. (1997a, 1997b and 1998). For him, any regenerated professionalism necessarily involves a strong emphasis on collegiality (amongst the members of the profession), negotiation (between professionals and clients), co-operation and collaboration (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 different values and outlooks'. Bottery (1998) covers similar ground, but provides a rather different account of what the ethical components of 'communicative' professionalism might be; he suggests they should include an appreciation of the provisionality of all knowledge, a commitment to truth seeking, reflective integrity and humility and a commitment to client empowerment. These, he argues, are 'the values needed to encourage the development of a strong democracy'. Communication then is central, but so too is the notion that the communication must serve purposes going well beyond the narrow interests of the professional; there is a strong value commitment to social improvement, the common good and the role of the public intellectual.

The second 'emergent' conceptualisation is of professionalism as reflective practice. One reason for the decline in respect for traditional professionalism was the re-theorising (perhaps re-conceptualisation would be a better term) of the nature of specialised knowledge. The seminal influence here was the work of Schön (1983 and 1987). The skeleton of his argument is as follows.

- (a) Professions have been honoured for their claim to extraordinary knowledge in matters of social importance and in return have been granted unusual rights and privileges.
- (b) That extraordinary knowledge (i.e. the body or bodies of specialised knowledge which could only be gained through extensive higher education) has increasingly proved inadequate, throwing the professions into a crisis of confidence and legitimacy.
- (c) The dominant epistemology of professional practice has been technical rationality; according to this view professional activity consists of instrumental

problem-solving made rigorous by the application of specialised knowledge and 'scientific technique'. (For a devastating critique of technical rationality in professionalism see Dunne, 1993.)

(d) The situations of practice are not however problems to be solved, but problematic situations characterised by uncertainty, disorder and indeterminacy. Technical rationality is inadequate in all those situations where problem setting (definition and framing of the problems) is at least as important as problem solving.

(e) Technical rationality may have some limited place, but should be subsumed within the broader concepts of reflection-in-action and reflective inquiry. To regain their place in society, professions need to increase their legitimacy through reflection-in-action and to encourage its wider, deeper and more rigorous use.

In recent years, a huge body of writing has been produced in which Schön's thinking has been highly influential. This is not the place to review that literature, but three critical comments on it seem necessary. Firstly, there is the routine fate of all influential writers: some of their disciples trivialise the message. It is depressing that many (particularly in teacher and nurse education) seem to think that all Schön said was that to be a professional you should just think about what you are doing. Secondly, evangelical persuasion of others to put the concept of reflection-in-action centre-stage, has resulted in an under-emphasis on reflection-for-action (before some episode of practice) and reflection-on-action (after such an episode). Thirdly, there is an epistemological objection to reflecting with knowledge-in-action which people are happy to let remain tacit and covert. The ability to reflect in, for, or on action still has to be nourished by bodies of knowledge which can, in principle at least, be made public. It can be agreed that specialised knowledge will always remain insufficient; this does not make it unnecessary and thus redundant.

There is a message within Schön's writing which, in this writer's opinion, has been unfortunately neglected or under-utilised, one which is of relevance to this research. For Schön the hallmark of the genuine professional lay in his or her capacity and willingness to engage in 'reciprocal reflection-in-action'. Reciprocal with whom? With their clients, their colleagues in the same profession, with members of other professions, with those who occupy key roles in the public policy process – but crucially also with those they have a responsibility to teach, train and develop. Perhaps then a key test of academic professionalism is whether or not academic staff can transcend patterns of expectation and liberate their students from limited learning by means of reciprocal reflection-in-action. If reflection turns experience into learning (Boud et al, 1985), then it can be argued that reciprocal reflection-in-action is the desirable process that produces meta-cognition and learning-to-learn. Here the concept of professionalism as reflective practice overlaps with communicative/interactive professionalism.

If Schön and his followers emphasise that professionalism has to become reflective, there are other writers who stress that professionalism must be critical (the third emerging conceptualisation). The main arguments for critical academic professionalism are cogently presented by Barnett (1997); useful examples of 'doing' critical professionalism are to be found in Walker (2001). Barnett starts from the main purpose of higher education which he defines as the development in students of 'critical being'. Critical being operates in three domains: knowledge, self and the world. Each domain attaches to a specific form of criticality: critical reason, critical self-reflection and critical action. Each form of criticality has different levels: for example, within the knowledge domain, critical reason can exist at the level of critical

skills (which are discipline-specific), of critical reflexivity (which involves critical thinking about one's own understanding), of critical thought (directed at re-fashioning traditions of thought) and of transforming critique (in which forms of knowledge are themselves re-formed).

Barnett's conception of academic professionalism derives from this analysis. If the academic professional is to be capable of supporting students in their growth towards critical being, then their own professionalism has to be re-conceptualised in a way which draws upon the triple schemata of critical reason, critical self and critical action. Such a notion of professionalism will centrally require capability in the critical deployment of multiple discourses. A fully-fledged professional will be adept in engaging not only with his or her students, but with a range of audiences, professional peers, members of related professions, managers and commercial stakeholders; s/he will have a duty to speak out to inform the public domain and cannot restrict professionalism to the arena of professional-client transactions alone (again merging into communicative professionalism).

Barnett might agree with Schön that professionals must necessarily be reflective practitioners, but would dismiss this as no more than a rather obvious rejection of professionalism as the simple application of specialised knowledge. For Barnett, the Schön conception of professionalism is unduly individualistic, crucially neglecting the extent to which professionalism must be social and inter-subjective; it also seriously underplays the theoretical components of critical reason in the knowledge domain and the importance of critical action.

Walker and her Glasgow colleagues adopted the Barnett analysis, but put great emphasis on the extent to which they believed that all academic professionalism must be oppositional and transformative; they consciously strove in their work for a

professionalism which would clearly be an alternative to traditional, diminished and enfeebled forms of professionalism which sought 'not to rock the boat'. As noted earlier, Walker and her colleagues had no wish to abolish professionalism, but rather to contest it and re-work what it meant. They quote Said (1993) with approval as defining the kind of professional person each of them was striving to become: '...a person motivated by care and affection who considers that to be a thinking and concerned member of a society, one is entitled to raise moral issues at the very heart of even the most technical and professionalised activity'.

The present writer would like to be able to offer a crisp re-conceptualisation to replace the four conceptions just outlined, some new synthesis. It would be helpful if such a synthesis could now emerge. This is an act of midwifery he cannot perform, for two reasons. Firstly, the conceptualising of professionalism has not proceeded by a series of dialectical swings from thesis to antithesis; thus to expect some synthesis to act as the new thesis is unreasonable. Secondly, professionalism is a complex, compound concept in which some elements have withered over time, some have been re-defined and into which new elements have been introduced. Any individual element may, in Raymond Williams' terms, be seen as residual, dominant or emergent, but it does not follow that the various conceptions of professionalism can be thus simply categorised; there is no one emerging form which is or should become dominant.

In this situation there might be thought only two ways of proceeding.

- (a) 'A range of meanings can be attached to the term professionalism. Here is a stipulative definition for the purposes of this research.'

(b) 'The participants in this research will attach their own meanings to the term professionalism. The only concern is with their conceptions.'

There is a third way and this will be adopted. This seems justifiable given the research approach and its underlying theoretical perspective. A personal, temporary, stipulative definition of what is to count as professionalism will be provided. This will guide the research, but may be fleshed out, modified or abandoned in the light of data provided by research participants.

For this research then, **professionalism is a quality required of people contractually engaged in work of serious public concern which implies the possession of (a) critical cognitive capability and (b) a commitment to independent, social and ethical action.** This requires elaboration. Firstly, it avoids the distinction between paid and unpaid work and substitutes for it an obligation implying accountability. Secondly, whether or not a public concern is 'serious' is allowed to remain problematic. Thirdly, critical cognitive capability embraces a number of requirements and criteria. The professional is required not only to possess those knowledge bases that are relevant but to be critical of them. S/he is expected to be aware of both the power and limitations of the understanding possessed. Critical reflection is needed on both the relative efficacy of means and on the desirability of ends. Capability is preferred to competence, both because it gives greater weight to underpinning understanding and because it goes beyond defining present skill requirements in terms of past functions to emphasising ability to cope with the future. Fourthly, the professional has to be committed both to independent action (where s/he accepts the responsibility for discretionary decisions in complex areas) and to social action, communicating with peers, other professionals and stakeholders; this

obviously requires judgement as to which form of action is appropriate. In addition an explicit commitment to ethical behaviour is necessary.

SECONDARY CONCEPT 1: CHANGE AND RESPONSE TO CHANGE

Change is an act or instance both of becoming different and of making different. For example, in the human sphere, a person can both change in the sense of becoming different (or being changed) and change in the sense of making different (of causing change). Individuals are both the objects of change and its agents. These two senses of the idea of change are important for this thesis. They underlie the intentional ambiguity in the title 'Changing assessment in higher education...' Assessment practice may be construed as there to be changed; it could be made different, perhaps by policy but also because of other change factors. Assessment practice may also be seen as something which is itself changing (becoming different for whatever reason) and which then has an impact on some other areas of academic life in higher education, including policy activity.

Fullan (1991) is perhaps the most influential author amongst the many who have been concerned with educational change in the context of institutional improvement. His central message is that 'real' educational change will only be achieved through pursuing both individual and institutional renewal. Individuals must take responsibility for empowering themselves through becoming experts in the change process; they have to engage collectively in continuous initiative to pre-empt the imposition of change from outside. Institutions will have to provide both pressure on, and support for, individuals. Policy-makers may have organisational power (principally through their role in resource allocation) but not educational power; they

can require change, but the changes which actually occur may not add up to educational progress.

During the development of this argument, Fullan makes a number of important distinctions. Firstly, 'symbolic' change is espoused to appease those exerting pressure from outside, to appear innovative or to gain additional resources, whereas 'real' change is sought because of educational commitments and values. Secondly, first-order changes are those aimed at improving the efficiency and effectiveness of existing practice; second-order changes seek to alter the fundamental ways in which institutions work, including new goals, new structures and new roles. Thirdly, the subjective meanings of educational change ('the multiple phenomenologies of the different role-incumbents') are very different from the 'objective realities' of change. This is not to say that the subjective realities ought to define what **should** change, but only that they are either powerful constraints on change or protection against undesirable and thoughtless change. Fourthly, within educational change three dimensions can be distinguished: the use of new materials, resources and technologies; the adoption of new approaches (new teaching strategies and learning activities) and the alteration of beliefs (underlying pedagogical assumptions and theories). For Fullan, change in all three dimensions is required for mere innovation to become educational progress.

Before leaving this author, there is one further argument that is illuminating. Fullan puts the strong emphasis he does on the **meaning** of educational change because to say that meaning matters is to say that people matter; educational change works, or doesn't work, on the basis of individual and collective responses to that change and these responses depend on shared meanings, shared cognition and what

Fullan calls 'interactive professionalism'. Such professionalism would 'go a long way in making significant change a reality'. (Fullan, 1991 – p 46)

Trowler (1998) focuses more specifically on how members of staff in higher education respond to changes imposed upon them by policy. Trowler's position is not easy to summarise, partly because his writing obfuscates more than it illuminates. He argues for 'a more complex model [of ways of responding] which incorporates the influence of not only normative elements and codes of signification, the rule-giving aspects of structure, but the authoritative and allocative aspects also'. The gist of his argument is that social and cultural structures have received too much attention in higher education research to date; such research has seen individual academics as merely responding to forces over which they have no influence. Trowler claims that his own re-conceptualisation of academic response allows a better understanding of the interplay between individual action and structural constraint.

The model of ways of responding then offered by Trowler is open to criticism. (Holroyd, 1999 a) Although some of Trowler's research participants were only accommodated within his model by Procrustean methods, nonetheless the model has sufficient credibility and utility to merit attention.

In the Trowler model, the response of practitioners to imposed policy changes depends both on how policy affects what they actually **do** and on how they **feel** about what they are required to do. Academic staff can be distinguished by how they position themselves on two axes; the first depends on the extent to which they generally like/approve of the central thrust of the policy change and the second depends on whether or not they view policy as something to be directly implemented or something that can be changed (or worked around). If a policy is liked and accepted, people swim; they go with the current, they flourish; they may exploit the

change for career advancement; they persuade others it is on balance benign. If it is disliked and accepted, they sink; morale and job-satisfaction decrease; there is a tendency to depression and people spend more time considering transfer or resignation. If a policy is broadly liked, but regarded as changeable, then that policy is actively reconstructed; people work harder to make the policy work better. If a policy is disliked, but regarded as mutable, then people respond with a range of coping strategies including policy avoidance and policy-demand reduction.

The emphasis on individual reactions to policy change is welcome, but there are reactions to policy change that the Trowler model does not seem to cover, because he attaches too little importance to workload and to resources. Two examples illustrate this. A member of academic staff faced with a policy change which is basically liked and approved may not embrace it if it is perceived as adding to an already heavy workload or if it has resource implications unlikely to attract extra funding. Some who dislike a policy they see as having to accept, but who are prompted to negotiate a change in workload, may well not sink. There are thus responses to policy change that have not occurred to Trowler; these include an increased commitment to policy activity (in the hope of prompting further change in policy) and an increased determination to alter the conditions of practice (to allow implementation of policy).

SECONDARY CONCEPT 2: DISCIPLINARY DIFFERENCE

From the start it was envisaged that the thinking of academics about assessment, about changes in assessment policy and practice, would be not wholly individualistic but influenced by the subject area in which they worked. An attempt to understand academics' perspectives on change and their responses to it should thus be

informed by an understanding of the nature of the disciplines in which they specialise. This line of argument, sometimes called ‘epistemological essentialism’, is central to the writings of, amongst others, Burton Clark.

As knowledge is newly created by research, and is reformulated and repeatedly transmitted in teaching, its face continuously bubbles up within daily operations, right in the palm of the professional hand. The logic, the identity, the very rationality of the academic profession is thereby rooted in the evolving organisation of those categories of knowledge that disciplines and professional fields of study have established historically and carried to the present, producing an inertia that powerfully prefigures the future.

(Clark, 1987 – p268)

Thus disciplinary difference was early recognised as a concept that should enter into the research planning; for example, the subject-area research sites to be selected should be as different as possible from one another. Operationalising this concept began on the basis of little more than a hunch formed by personal experience and then, in a pleasingly fortuitous way, relevant literature came to light. Rather than deal with that literature here, consideration of it will be postponed; it appears as ‘Interpolation – a Theoretical Model’ within Chapter 3. This unorthodox approach is truer to the chronology of the research process.

RELATIONSHIPS BETWEEN CONCEPTS

There are three primary concepts: policy, practice and professionalism. There are two pervasive, secondary concepts: change and disciplinary difference. If there are relationships between these concepts then we have the beginnings of a theoretical framework. At this stage it is appropriate to indicate the notions about possible relationships which were entertained in the early stages of the research. Later (Chapter 10) there is discussion of how the framework developed because of the

research. The simplest way of indicating the primary concepts and the suggested relationships is as below.

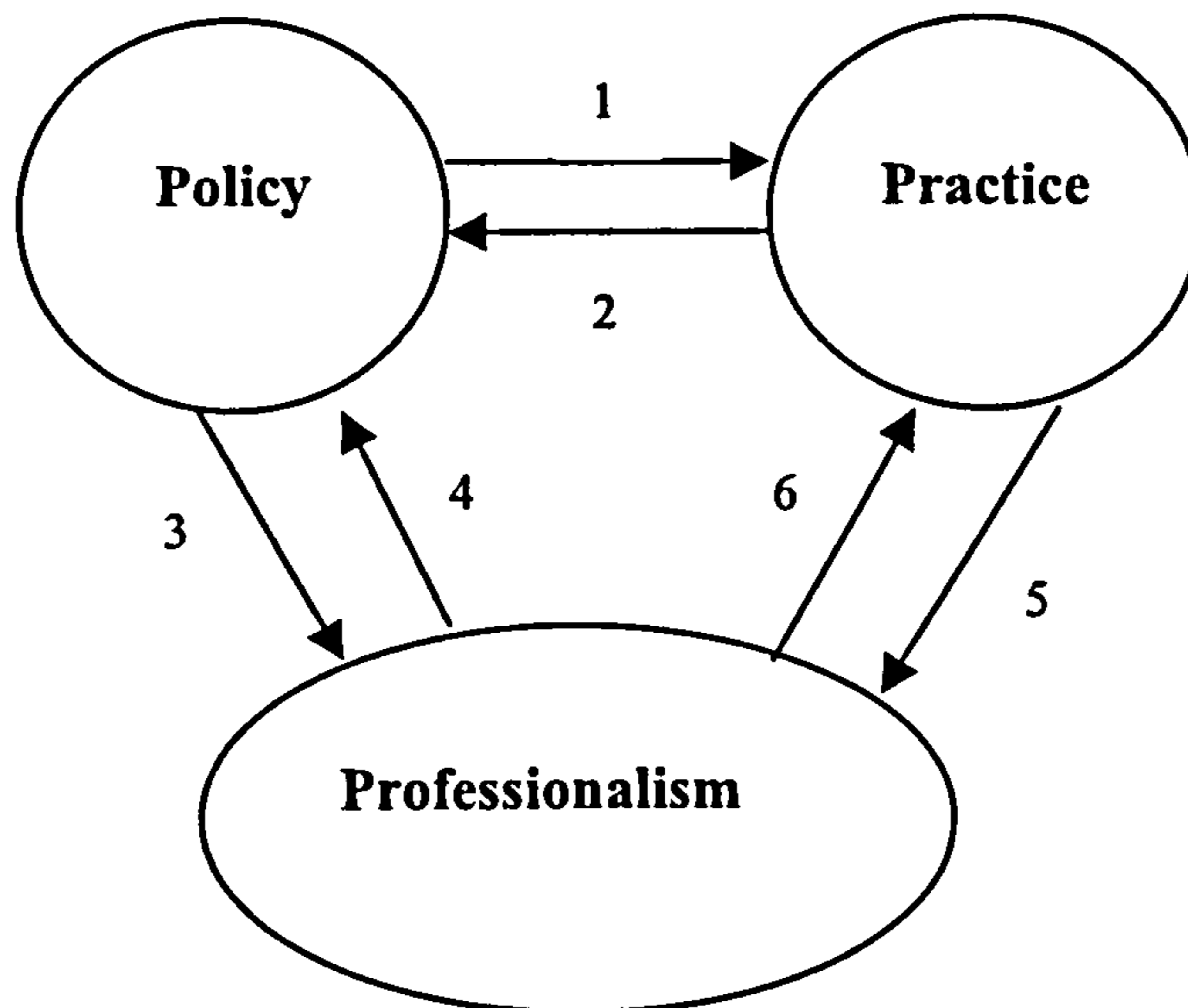


Fig. 2 Relationships between primary concepts

The ‘early notions’ of the links have the status of common-sense, plausible assumptions. They are the kinds of links that might readily spring to mind if assessment practitioners were asked ‘How, simply, do you see X and Y as likely to be related?’ Some evidence that they are not merely the idiosyncratic ideas of the writer can be found in Appendix XI. That these suggested links are too simple is already clear from the exploration of the concepts in this chapter.

1. Assessment policy is an important cause of change in assessment practice.

If policy is seen as embodying a proposed course of action, then policy makers assume that the action they propose will in fact be put into practice. The formulated policy causes existing practice to change. Looked at from the practitioners’ point of

view, new policy relating to their assessment practice emerges and has to be implemented; changes in assessment practice are then the effect of the new policy.

The statement suggests that assessment policy is only one cause, an important one, of change in practice. In other words, it is being assumed that if practitioners are asked to explain changes in their assessment practice, they will identify policy as one influential factor amongst others which has caused practice to change.

2. Existing assessment practice causes the emergence of new policy.

There can be elements of existing practice that those in authority may judge to be in need of change, so they make policy to achieve this. But practice need not be placed in this deficit frame; some positive development in practice may raise matters that existing policy does not cover and thus prompt a revision of policy. In both cases practice causes policy, or new policy is the effect of existing practice. This relationship could be at departmental/faculty level: 'What we do is such that we should generate policy moving practice nearer to what it ought to be'. It could also be at institutional level: 'What is going on in various departments/faculties is such that we need to formulate policy to regulate it'.

3. Emerging assessment policy causes a restriction of professionalism.

Policy is authorised decision-making and comes from the top, down through hierarchical levels to the practitioner. However, traditional professionalism assumes that the practitioner has been entrusted with discretionary powers in the making of difficult decisions in complex areas like assessment. Obedience to the dictates of policy may restrict the independence and autonomy of the professional. i.e. it may de-professionalise him or her. It does seem necessary, however, even at this early stage,

to distinguish between policy which causes a change in how professionalism is operationalised and policy which prompts a re-thinking of the nature of professionalism itself i.e. how it is conceptualised.

4. The greater the professionalism of academics, the greater their input to policy.

The assumption here is that claiming or possessing professionalism implies a recognition that one's activity is not a wholly private matter and that on occasion one will be obliged to contribute to the policy context in which one operates. For example, the social and ethical values embedded in professionalism can imply an obligation to uphold and promote the democratic values of civilised society. The academic professional is thus seen as having a duty as a public intellectual to make a contribution to the policy-determining process. A professional capability in critical thought and a commitment to critical action should be put at the service of the common weal. In a more limited sense, the greater the professionalism in assessment, the more active the professional will be in trying to influence the policy context and the conditions in which assessment practice is conducted.

5. The more change in assessment practice, the greater the demands on professionalism.

Even if assessment practice continues unchanged for lengthy periods, there is still a requirement for some professionalism; the practitioner is under a professional obligation to continue to do 'good work'. The assumption being made in this statement is that when assessment change is being contemplated or when some new assessment procedure is being implemented, the demands on professionalism are

increased. This arises in several different ways. If an individual wishes to introduce some innovation in assessment, s/he has to reflect more on existing practice and on the possible alternatives; s/he may be required to acquire new knowledge and understanding. When policy requires some change to be made, there is usually the need for practitioners to consider how best the change is to be implemented in the interests of their particular subject and their students. The effective introduction of changed assessment procedures and systems usually requires more collaboration and co-operation, sometimes with existing colleagues and frequently with new partners. To assess conscientiously and wisely is never easy; to continue to do it well using new methods in different conditions is likely to prompt practitioners to question the adequacy of their existing professionalism or perhaps to re-define what is to count as professional.

6. Where there is professionalism, there is 'good work' in assessment practice.

The claim to professionalism carries an obligation to do good work. If assessment is carried out in invalid, unreliable and unfair ways, then the reasonable judgement of peers, of clients and of the public will be that the practitioner has acted unprofessionally. 'Good work' is expected of the professional and should be possible within the existing systems and conditions, but if it is not then there is a professional pressure to change assessment practice in ways which make 'good work' possible. Professionalism requires the practitioner both to be competent within the existing system of practice and, being critical of it, creatively to change it.

This introduction to 'the links' has involved the pervasive concept of change, but no mention has been made here of disciplinary difference. The initial assumptions were that there would be subject-related differences (a) in assessment thinking and practice, (b) in perceptions of policy and response to policy change and (c) in how professionalism was construed. There was a 'hunch-hypothesis' that the four-quadrant framework would prove useful in throwing light on these differences.

THE NATIONAL POLICY CONTEXT

The Dearing Report

The assessment of student learning is changing, and will continue to change, partly in response to recent and continuing emergence of national policy on assessment. Although it would not be true to suggest that such policy has only been in existence over the last five or so years, it is only since the report of the National Committee of Inquiry into Higher Education (NCIHE, 1997) – the Dearing Report – that national policy on assessment has achieved a high profile. The Committee made recommendations to Government and, of course, all the recommendations did not automatically become official national policy, as the issue of student fees made very clear. Nevertheless, in general terms, the Dearing recommendations which relate to assessment remain intact as policy requirements and will continue to have some significant impact on assessment practice.

Is there a paradox? The Dearing Report attaches a high priority to the promotion of learning and to the enhancing of standards of student achievement, but in its 88 main recommendations there is no explicit mention of assessment at all. (The word 'assessment' is used only in relation to the Research Assessment Exercise in recommendations 32, 33 and 34.) The full report (NCIHE, 1997, Vol 1) does refer to

student assessment and directly addresses the topic (on pages 137-139) – but this section has no recommendations. Does this then demolish the earlier claim that the report will have a significant effect on assessment practice? The paradox disappears if Dearing saw assessment as part of any effective strategy for the management of learning, one that was so obvious that explicit reference was unnecessary.

Taking this view, then several of the recommendations are important for 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 (to cover external examining and student assessment) to be adopted by all universities as a condition of continued public funding (recommendation 24) and the setting-up of small expert teams to provide benchmark standards/statements (recommendation 25). These recommendations deserve more extended analysis than is possible here (see, for example, Smith et al, 1999). A quick comment is necessary on benchmarking and the Code is the subject of the next section.

Benchmark statements 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 statements extends. Firstly, they apply to honours degrees but they could 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 seems to be 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 matter 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, when this prescription is taken along with the exhortation that assessment procedures be valid, there is a strong pressure for methods of assessment to be scrutinised and where necessary changed.

The QAA Code on Assessment (April 2000)

The Code of Practice for the Assessment of Students in UK Higher Education Institutions (QAA, 2000) is one of a suite of inter-related documents which together form an overall Code of Practice for the Assurance of Academic Quality and Standards. Each section of the code is structured into a series of precepts; each precept is accompanied by outline guidance. The precepts indicate key matters that QAA expects each institution to be able to demonstrate it is addressing effectively through its internal quality assurance procedures. The guidance is ‘not intended to be either prescriptive or exhaustive... nonetheless, in every institution the guidance will constitute appropriate good practice’.

The word ‘precept’ has not occurred in this thesis so far. In the absence of clarification from QAA, it is assumed to have its standard meaning of ‘command’ or ‘rule of conduct’. Given this meaning, it is a little odd that seventeen of the eighteen precepts on assessment contain the word ‘should’ rather than ‘must’; the eighteenth contains ‘will normally’. Allowing that the distinction between what is educational and what is administrative is not always as clear as it may seem, then only three of the eighteen QAA precepts are educational i.e. concerned with **how** student learning

should be assessed. These three are as follows. Number 2 The principles, procedures and processes of all assessment should be explicit, valid and reliable. Number 6 Institutions should ensure that... assessment is consistent with an effective and appropriate measurement of the achievement by students of the intended learning outcomes and effectively supports learning. Number 12 Institutions should ensure that appropriate feedback is provided to students on assessed work in a way that promotes learning and facilitates improvement.

The Code prescribes national policy on assessment which a university can not ignore. The bulk of the Code is concerned with the administration of assessment. The educational element is confined to three areas (emphasising explicitness/ validity/ reliability, constructive alignment and helpful feedback). In the first of these both principles and processes/procedures to implement them are covered; in the second and third only the principle is stated, not the ways in which it should be operationalised.

A European Dimension?

There is an assumption in the heading of this section that the only supra-institutional policy affecting assessment is national. In future it may well be supra-national as well. For example, in 1999 the Education Ministers of the European Union signed the Bologna Declaration which deals with the adoption of comparable degrees as a contribution to the economic development of the Union. (BOLOGNA, 1999) This connects with assessment, dealing as it does with students on placements abroad, international exchanges, distance-learning programmes and general issues of compatibility of grading schemes and comparability of standards. The Bologna Declaration appears to have had relatively little impact on UK higher education so far, but it is predicted to become much more significant as this decade progresses and the deadline for implementation approaches. (Yorke, 2001)

CHAPTER 3 THE RESEARCH ACTIVITY

ASPECTS OF PLANNING (1): THE PARTICIPANTS

In the research activity of the minor (policy) strand there was a sense in which all members of the policy-generating group (The Assessment Working Group) including the researcher, were participants. The two people interviewed were the Vice-Principal responsible for setting up the group and the Convener of the group.

For the major (practice) strand, about eight interviews were considered manageable in each subject site. These eight would not be a representative sample of all academic staff in the subject area; rather they would form a purposive sample (Cohen et al, 2000, pp 103-4) chosen because their experience clearly qualified them to talk about the topics in the interview. The intention was to seek people as follows:

- (a) the head of department (or a senior member of staff nominated by him/her)
- (b) the chair of the learning and teaching committee;
- (c) the class co-ordinators for Level 1 and Honours courses;
- (d) the examinations/assessment officer;
- (e) three members of staff with a significant workload in assessment but no particular management role; one with long experience of the University, one with experience of assessment in another university and one 'junior' member (a graduate teaching assistant or a probationary lecturer).

The decisions about research sites and people within them were recognised from the planning stage to have clear implications for the type of generalisation that

would be possible from the cases studied. There would be no claim that the people interviewed were representative of all academic staff in that area, nor that the four subject areas made up a sample representative of the University as a whole. There could thus be no empirical generalisations from those interviewed to all staff in that area, or from the four subject areas to the University as a whole, or indeed from the University to all other universities. Naturalistic generalisation and theoretical inferencing would be legitimate forms of generalisation. (OU, 1996, Chapter 3.1)

THEORY: AN INTERPOLATION

In devising a theoretical framework to guide the choice of subject sites, the writer thought he had been genuinely creative. He then read, in 2000, *Academic Tribes and Territories* (Becher, 1989); this reading was accidental and fortuitous. In that book (Chapter 2) Becher outlined a framework for categorising academic territories that was in essential respects the same. (See also, Becher and Trowler, 2001)

It may be helpful to outline how Becher arrived at his theoretical framework. Taxonomies of knowledge fields from Pantin (1968) and Kuhn (1977) were uni-dimensional and proposed simple two-fold categorisations, in the former case of knowledge forms as either restricted or unrestricted and in the latter of their being in paradigmatic or pre-paradigmatic stages of development. These taxonomies were based on detached observation of how researchers operated within their different domains. More phenomenological analyses, i.e. ones based on how the academic actors, staff or students, themselves perceived the areas in which they were engaged, came from Biglan (1973) and Kolb (1981). Biglan suggested three dimensions: hard versus soft (equivalent to high and low degrees of paradigmatic development), pure

versus applied and life systems versus non-life systems. Kolb's analysis derived not from staff perceptions but from the learning strategies adopted by students in different subject areas. His was a two-dimensional framework: one axis covered abstract-concrete variation (contrasting a bias towards conceptualisation with a domination by immediate experience) and the latter covered active-reflective variation (contrasting a preference for active experimentation over detached observation). Kolb was able to show (when the life versus non-life systems distinction was dropped because it showed little variance) a 'high consistency' between how staff saw their subjects and how students operated within them, all the more remarkable given that the research approaches were so very different. Kolb concluded as follows.

The commonly accepted division of academic fields into two camps, the scientific and artistic, or abstract and concrete... might be usefully enriched by the addition of a second dimension, namely active-reflective or applied-basic. When academic fields are mapped onto this two-dimensional space a four-fold typology of disciplines emerges. In the abstract-reflective [hard pure] quadrant are clustered the *natural sciences and mathematics*, while the abstract-active [hard applied] quadrant includes the *science-based professions*, most notably the engineering fields. The concrete-active [soft applied] quadrant encompasses what might be called *the social professions*, such as education, social work and law. The concrete-reflective quadrant [soft pure] includes the *humanities and the social sciences*.

(Kolb, 1981, original emphasis)

Becher decided that a Kolb-Biglan classification of academic knowledge was a particularly appropriate one for his purposes, though preferring Biglan's more accessible contrasts between hard and soft, pure and applied to the esoteric terminology used by Kolb. The reasons were, firstly, that such a framework was capable of more subtle distinctions than the uni-dimensional scales proposed by Kuhn and Pantin among others, but was not so complex as to become unfunctional

and, secondly, it was more directly related to established academic groupings than the well-known abstract categorisations of say Hirst (1974) and thus directed attention not only to the epistemological properties of knowledge fields but also to the social classifications of academic groups (i.e. to both territories and tribes).

In this present research, the choice of subject areas was guided by a Holroyd-Becher-Kolb-Biglan framework. As already stated, it then became an additional purpose of the research to discover whether interviewees located their subjects as would be predicted from the theory. This left undetermined how the framework was to be presented to the research subjects/partners. Although it was adopted as a general classification of knowledge fields to guide choice of subject sites, within the research it was to operate as a way of distinguishing how staff thought about assessment within their fields. In other words the terms pure, applied, hard and soft had to be translated into simple and meaningful descriptions of different perspectives on assessment.

This translation was assisted by the preliminary discussions with people from different subject areas. It was encouraging that they seemed to accept the framework as a whole as very reasonable. (In addition several made the point that focusing on assessment was appropriate, because people ‘dressed-up’ their subjects in what they wrote in course descriptions and in what they said to students but revealed their ‘true’ thinking in what they selected to assess.) In these discussions there was uncertainty about how best to capture the essence of the dimensions in a few words. The researcher came to the conclusion that the axes should be presented as ‘ways of describing what gets emphasised in your assessments towards the end of degree programmes’. ‘Pure’ was described as ‘assessing students’ academic soundness’; ‘applied’ was described as ‘assessing their fitness to practise’; ‘hard’

should be taken to mean that ‘there is a bundle of core content that students are expected to possess’; ‘soft’ should be taken to mean that ‘what gets assessed is student ability to evaluate alternatives - arguments, products or whatever’. Those interviewed found these elaborations, on the whole, meaningful and acceptable.

ASPECTS OF PLANNING (2): RESEARCH APPROACH AND METHODS

The research purposes were defined by the research aims and questions given in Chapter 1. Research methods were chosen on the basis of judgements about their fitness for purpose i.e. in terms of their ability to deliver data which would contribute to answers with some degree of credibility.

This is too simple, for two reasons. Firstly, the idea that research planning moves along a simple linear path from purposes to methods is part of some abstract model that bears little relationship to the ‘real’ process of design. The definition of purposes, the selection of approach and methods and the choice of research all either proceed in parallel or are continually revisited in the iterative process of design. Secondly, the research questions given in Chapter 1 evolved from earlier versions; that evolution was influenced by the researcher’s view that some research approaches have a more satisfying rationale than others and also by personal preferences for some research methods over others. Research questions carry within them implications for the choice of appropriate methods; preferred methods of research influence the way in which research questions are expressed. There is thus an obligation to make explicit the kind of research that this writer wished to carry out. This can be summarised in six brief statements (see also p. 13); the specialist terms used have the meanings attached to them by Crotty (1998).

- (a) The epistemology underpinning the theoretical perspective of the research would be constructionist rather than objectivist or subjectivist.
- (b) The theoretical perspective would be interpretivist rather than positivist or post-modernist. The form of interpretivism would be symbolic interactionism
- (c) The approach (or methodology) would be ethnographic, liberally construed.
- (d) There would be no artificial distinction between qualitative and quantitative research; there is no theoretical prohibition to using both within one research project.
- (e) Semi-structured interviewing, participant observation and documentary analysis would be more congenial (for this researcher at this time) than, say, survey by questionnaire or systematic observation.

This writer has been amused by the summary self-labelling so beloved of some researchers: 'I am of course a post-structuralist...' or 'My research approach is Foucauldian...' If pressed to label himself he would wish to be known as an eclectic post-positivist and a subtle realist. Some justification for these labels, and for the six statements in the previous paragraph will be found in Appendix III.

The principal method of research chosen was interviewing, predominantly semi-structured but with two structured episodes included. There was also some participant observation and documentary analysis. The researcher had some experience of interviewing (Harlen and Holroyd, 1995). Nevertheless, to update his knowledge and improve his understanding of the method he read (and re-read) widely. The most helpful texts were those from Powney and Watts (1987), Fontana and Frey (1994), Drever (1995), The Open University (OU, 1996), Kvale (1996),

Bassey, M (1999), Verma and Mallick (1999, Chapter 6), and Cohen et al (2000, Chapter 15). Inside the front cover of his research file he typed the following in large print.

- AIMING FOR.....***
- (a) Short questions prompting long answers.***
 - (b) Probing to clarify relevant meanings.***
 - (c) Interpretation throughout.***
 - (d) Verification of hunch interpretations.***
 - (e) Intelligent assertiveness with emotional warmth.***

Some authors include methods of data preparation and analysis under the general heading of research methods. If this is accepted, then methods employed were data reduction, content analysis and thematic analysis.

WHAT RESEARCH WAS CARRIED OUT? WHEN?

The minor strand: The Assessment Working Group

The author was an ordinary, fully-participating, member of this group (which was intended to generate policy) from its setting-up in January 1998 until June 1998. When it was realised that the work of the group could provide a useful strand for the planned research, he sought continued membership as both participant and observer; the Convener and members agreed to this. He then attended every meeting, except one, as a participant observer from September 1998 until January 2002. His role as researcher was never referred to after the initial negotiation and it seems unlikely that his presence as a researcher had any significant effect on the workings of the group. The research obligations did, however, mean that he was able to contribute less frequently during meetings and there were occasions when he

was aware of role-tension, if not conflict. For example, when asked to take his share of drafting sections of an assessment code he readily agreed; when asked to be the sole author of a guide to good practice in assessment he felt this incompatible with a research role and refused.

At all meetings of the group, copious notes were taken; as this represented only a minor change in his normal behaviour, it aroused no comment. After each meeting, brief 'field-notes' were written-up and appended to the notes made during the meeting. These helped to ensure the notes would remain intelligible; they identified questions that seemed to be worth pursuing and insights that might influence progressive focusing of attention during subsequent meetings. Agendas and minutes of meetings were collected, as were all documents and reports provided to the group and all drafts of papers produced by members of the group.

During one phase of the committee's workings, an attempt was made to build on a form of episodic analysis devised by the writer during earlier research into the business of school boards. (Munn and Holroyd, 1989) This allowed a crude quantitative estimate of the time devoted by the group to different aspects of its work. For example, one-sixth of the total time was devoted to discussion of how the remit should be interpreted and fulfilled.

When it became clear that the group would not finish its work by the target date and would be expected to go on meeting, perhaps for years, and when the piles of paper became overwhelming, the researcher decided that the best he could do was to continue with the conscientious accumulation and cataloguing of material and postpone all serious analysis to a later date. By June 2001, it was obvious that sufficient data existed on which to base several research projects, not just a minor strand within one project. There was a strong temptation to jettison this whole

aspect of the research. Rather than do this, two decisions were taken. Firstly, the first two research questions were narrowed from their previous form to the versions given in Chapter 1. This allowed a drastic winnowing of the data; much could be rejected as irrelevant to the re-defined purpose. Secondly, the Vice-Principal responsible for setting up the group and its Convener would be interviewed; these would be semi-structured interviews based on an outline schedule shaped by the revised research questions. These two interviews duly took place in July 2001.

The major strand: preliminary discussions

Conversations were held with seven people. A ‘conversation framework’ was devised and notes were taken on this during each unrecorded conversation. The framework had four sections: the purposes and scope of the research; the choosing of subject area sites; appropriate people to interview within sites; diplomatic and ethical aspects of approval and anonymity. The seven were the Vice-Principal (Learning and Teaching), the Director of the Teaching and Learning Service (TLS), two members of TLS staff (one with research and one with assessment expertise), a senior lecturer in psychology (described as hostile to research in all paradigms other than his own), an Associate Dean well-versed in the ways of the University, and the former head of a large subject department.

The atmosphere of these 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 approval was freely given. A summary of the main points emerging from the conversations is given in Annex D to Appendix II. These conversations were completed by May 2000.

The Vice-Principal recommended the production of a written protocol on aspects of approval and anonymity ('the rules of engagement'). He revealed that he routinely required such a thing of any external person seeking to do research within the University and that 'this regularly stops them going ahead altogether'. A protocol was prepared, discussed and approved (Appendix XII); it was then adhered to throughout the research.

The Major Strand: Access to Sites

Negotiations for access to people in four subject areas were successfully completed by June 2000. Details of the process are to be found in Section 3.2.2 of Appendix II.

The Major Strand: Piloting of Interviews

A draft of a possible schedule was prepared and then piloted with three people (a) to improve the schedule and (b) to remind the researcher of his strengths and weaknesses as an interviewer. The people chosen were from some of the categories of people to be interviewed later, but they were not themselves interviewed again i.e. they did not form part of the final samples. The interviews were conducted as serious dry-runs and were recorded. Pilot interviewees all agreed to discuss the process at the end of the 'formal' interview.

The schedule was modified after each pilot interview. The version given in Appendix XIII was then used for the main interviews, with only minor variations in wording between one subject area and the next. A summary of salient points from the piloting phase is provided in Appendix II. The pilot interviews were completed by July 2000.

The Major Strand: Activity in the Four Subject Areas.

The work done in each area was roughly similar and went through the following main phases.

- (i) Agreement was reached with the liaison person nominated by the Head of Department or Dean as to the eight or so people who would form the purposive sample for each subject area, with one or two reserves should anyone be unwilling to participate. The liaison people were as follows: chemistry – the Chair of the Learning and Teaching Committee; philosophy – the Head of Department; Medicine – the Associate Dean (Education); design – the Chair of The School of Design and Craft. A significant difference emerged during this process of consultation and negotiation. In the first two ‘academic’ subjects, it was thought over-ambitious to try to cover assessment at all stages in the degree course; it was thus agreed that there would be a particular, but not exclusive, focus on the first and last years (i.e. Level 1 and Honours courses). In the third and fourth ‘vocational’ subjects, the liaison people were strongly of the view that their assessment systems were coherent and progressive throughout all years of the course and that equal attention should thus be given to all years.
- (ii) The liaison people described above advised on the formation of the four purposive samples of interviewees. The intended composition of the samples was as described on page 48. The actual samples from chemistry and philosophy departments were exactly as intended. In medicine and design the subject area was covered by a faculty and a school, rather than by a department, and different structures meant the samples had to be somewhat different. In medicine a sample of eleven was needed to cover the functions and characteristics of the eight in

chemistry and philosophy and one of these was a member of academic-related staff. (The functions of examination officer were carried out by the faculty clerk.) In design, to ensure coverage of specialist sections within the school, the sample size was increased to nine. Details of the sample composition for medicine and design are given in Appendix VI, page 2 and Appendix VII, page 2.

(iii) Potential interviewees were approached by letter in the first instance; this was followed up by e-mails and telephone calls. In three of the four areas, there was one person unwilling to take part and a replacement was found. In the fourth area everyone who was approached agreed. Convenient dates and times were easily and pleasantly negotiated.

(iv) Course documentation, particularly that with relevance for assessment, was collected both as hard-copy publications or leaflets and as downloaded material from the subject web-pages. Familiarity with these was gained before the interviews and proved especially useful in the section of the interview devoted to how students are currently assessed. The considerable time spent in studying this documentation was well-spent; interviewees welcomed the fact that their interviewer appeared to know what he was talking about. It was interesting too that in all of the subject areas some interviewees observed, without any prompting, that their course documents were not as clear as they should be in relation to some aspects of assessment.

(v) The interviews were carried out during the following periods.

Chemistry: eight interviews: September – October 2000

Philosophy: eight interviews: November – December 2000

Medicine: eleven interviews: January – February 2001

Design: nine interviews: March – April 2001

In all cases except one the participants readily agreed to the interviews being recorded. The average length of interviews was 54 minutes; interviews ranged in length from 38 to 75 minutes. In a dozen cases the interviews were followed up by e-mail interaction and/or the provision of additional printed material. All interviewees were thanked by letter immediately after the interviews.

At this point the writer faces a dilemma. What has been said about these interviews is bloodless and boring; these were rich interactions with complex and interesting human beings. How is their flavour to be conveyed within the constraints of available space and the conventions of research reporting? Five quick points must suffice.

(a) All thirty-six interviews took place without any diplomatic incidents or unpleasantness. This is a tribute to the courtesy of hard-working academics. Perhaps also it is some tribute to the inter-personal skills and administrative abilities of the researcher. It is also possible that he failed to confront matters that deserved to be confronted; within a few interviews there were places where a desirable intellectual assertiveness was over-ruled by the demands of imperturbable amiability.

(b) Most of the interviews were not only interesting but enjoyable for the researcher; there is substantial evidence that those interviewed **also** found the experience enjoyable.

Thank you for that. I didn't expect it to be enjoyable – but it was.

I have never talked about assessment for so long in my life. It's one of those topics which becomes more interesting the more you think about it.

- (c) All the interviewees addressed the interviewer's questions seriously. Even when assessment was perceived as an unwelcome burden added to the demands of teaching, administration and research, it was recognised as meriting careful attention because of its significance in the lives of students.
- (d) The tapes of the interviews were surprisingly often punctuated by laughter. Sometimes this followed something funny; more often than not it was a meaningless social noise indicative only of shared humanity.
- (e) There was evidence in about half of the interviews of an ambivalence in how the interviewees perceived the researcher. He was a junior researcher and thus it was appropriate to educate, enlighten and occasionally to patronise him; he was also a senior academic developer from whom it might be possible to learn something. Readers are invited to interpret the following four comments for themselves.

I'm not seeing you out of the building; I'm just going to the toilet.

I'll just have my mid-morning coffee while I talk to you. (At 9.00 am...)

We need someone like you to tell us what to do about assessment.

If you were testing me on assessment, would I have passed?

The Major Strand: Analysis and Reporting

The sequence adopted in this account does not mean that only after data collection was complete did the process of analysis begin. Some form of analysis was taking place right from the preliminary conversations with key informants, through the pilot interviews and during the main business of interviewing. Within the interviews, there was a continuous attempt to analyse what was being said and to modify future interjections in the light of that analysis and in the hope of making the interaction more productive. Between interviews there was some progressive

focusing, all done within the broad structure of the interview schedule. Nevertheless a more intensive phase of analysis did occur after each group of interviews was complete.

After the first set of (chemistry) interviews, three of the longest and, on an initial impression, the most 'useful' interviews were transcribed verbatim and in full. From these transcripts was induced a possible coding and categorisation scheme. That scheme was then tested and, where necessary, modified by listening again to the three tapes and studying the transcripts. A loose-leaf file was then prepared with one sheet for every sub-category, loose-leaf so that the sheets could be shuffled and supplemented. This 'analysis file' was then completed by listening to all eight tapes, stopping and re-playing as required and inserting data into the appropriate sheets with references to the interviewee identifying number and with cross-references when it seemed appropriate to insert the same data under more than one heading. Vivid or illuminating quotes were entered in full for possible future use in enlivening the reports. Prior to beginning this process, serious thought was given to using computer software for qualitative data analysis; on balance it was decided that 'hand-coding' was preferable (a) because the researcher did not wish to lose any of the benefits of familiarity with the data that come from regular immersion within them and (b) in a previous research project, involving the analysis of sixty lengthy interviews, the writer had learned that he could work up a reasonable speed in data coding and entry after the initial, necessarily very slow, phase. As a check on his coding, he enlisted the co-operation of a research mentee in three sessions when they worked together on excerpts from the interviews that seemed ambiguous or problematic. A similar coding/categorisation process was adopted in the other three areas.

This account has dealt inadequately with the nature of the analytical methods employed. In essence, two types of analysis were carried out making use of the coding/categorisation scheme outlined. Firstly, content analysis aimed to provide answers to the research questions posed; secondly, thematic analysis sought to identify those 'emerging' issues that were important to the interviewees and to develop those 'imported' issues that the researcher had already identified as relevant.

The researcher had suggested that staff in each subject area would welcome a report dealing only with their own subject; this had been endorsed by the liaison people. The writer committed himself to producing four separate subject area reports. This strategy made possible an exercise in informant validation. Each report was drafted and sent to all participants in the relevant area with a carefully-worded covering letter. Just less than half of the interviewees responded to this invitation. The researcher was left unsure whether he should be pleased (very few factual errors were identified, there were few differences of opinion on matters of interpretation and seemingly few lapses from clarity) or disappointed (perhaps some people did not bother to read the report at all or the report was not sufficiently interesting to find time to respond to). The impact of these reports is reviewed in Chapter 10.

It was decided to call these 'descriptive reports'. The title is a little misleading; a considerable amount of analysis was required to produce the reports and they often go beyond description. The title was adopted in an access of modesty and in the full awareness that further analysis would be necessary. After giving due thought to the responses of informants and making revisions where appropriate, a second version of the report was produced and sent to the departments or faculty.

The dates of distribution for the draft and revised versions were as under.

	Draft Report	Revised Report
Chemistry	November 2000	December 2000
Philosophy	February 2001	March 2001
Medicine	June 2001	August 2001
Design	September 2001	October 2001

Early in October 2001, the researcher met the Vice-Principal to report on progress and to lodge copies of the four reports. He pronounced the reports ‘fascinating and very useful – but too long for my purposes’ and requested a summary version. This report was delivered at the beginning of November 2001 and it also formed the basis of a research seminar that same month. The four subject reports and the summary report are provided as Appendices IV to VIII.

Further analysis of the data and integration of the material from the minor and major strands continued at intervals until the writing-up of this thesis.

CHAPTER 4 CHANGING ASSESSMENT POLICY

THE ASSESSMENT WORKING GROUP: ACTIVITIES AND REPORTS

Phase 1 January to June 1998

The Assessment Working Group (AWG) was set up in January 1998 with the remit 'to develop a strategy on assessment within the University for consideration by the Committee on Educational Development and Innovation, the Education Committee and the Senate'. In tackling their remit the group was required to take into account general University policy ('The Strategic Planning Statement') and the relevant recommendations from the Dearing Report and 'other national policy developments'. Also listed were twenty-two substantive assessment issues the AWG should address. The group was small: there were five 'ordinary' members (including the present writer), two ex-officio members (one Vice-Principal and the Clerk of Senate) and a clerk. The AWG met thirteen times and produced a 'Final Report' in May 1998.

Phase 2: September 1998 to June 1999

The Committee on Educational Development and Innovation (soon to be renamed the Committee for Educational Strategy and Resource) took the report of AWG and reports from two other working groups (Teaching and Learning; Key Skills, Careers and Work Placements) and produced a paper 'Towards a Learning and Teaching Strategy' for discussion in the Education Committee and in Senate (May 1999). This was the University's response to recommendation 8 from Dearing (NCIHE, 1997) that 'with immediate effect, all institutions of higher education give

high priority to developing learning and teaching strategies which focus on the promotion of students' learning'.

The AWG was re-animated in January 1999 with the same membership and met nine times before June of that year. Added to the remit of the group was that it produce a Code of Practice on Assessment, by June 1999. The AWG quickly decided that such a Code should contain eight main sections: (i) Assessment: Principles and Purposes, (ii) Designing an Assessment Scheme, (iii) Implementing an Assessment Scheme, (iv) Providing Feedback, (v) Standards, (vi) The Management of Assessment, (vii) The Assurance of Standards and (viii) Central Administration of Assessment. Each member took responsibility for drafting and developing at least one of these sections. During this period the AWG appointed a research assistant whose main task was to collect and analyse codes of assessment practice from other universities in the UK and beyond.

By June 1999, the AWG had made considerable progress with most sections of the Code, but had not managed to develop some parts (particularly Section (v) - Standards) in a way that satisfied it. The AWG had also decided that it could not fulfil its remit by the production of a Code alone; this was to be supplemented by a Guide to Good Assessment Practice. A draft of some sample sections of this was also available by June 1999.

Phase 3: September 1999 to June 2000

The AWG continued with the same members and met four times before producing a progress report (December 1999). Almost all of the time was taken up with the Standards section of the Code; the group found great difficulty in formulating generic class and level descriptors and procedures for aggregation. In

the progress report, the AWG regretted that it had failed to discharge its remit by the deadline 'but quite simply it has been overtaken by the magnitude of the task'.

There were a further three meetings before the group produced, in March 2000, a Draft Code of Assessment. The covering letter to this made it explicit that the draft was incomplete in two respects: firstly, the Code made provision for University-wide grade and honours class descriptors, but no draft text for these was provided and, secondly, the Code envisaged a single University-wide aggregation mechanism but no specific scheme was proposed. The letter also informed people that the Code would later be supplemented by 'a comprehensive guide which was still in the process of completion'. From March until June 1999, the draft code was out for consultation within faculties and within the parent committees of the AWG.

In February and March 2000, the Quality Assurance Agency issued for consultation draft versions of a code of practice for the assessment of students. The definitive version was published in April 2000.

Phase 4: September 2000 to June 2001

In October 2000, the Vice-Principal convened a special meeting attended by two representatives from every faculty and members of the AWG to consider faculty responses to the draft Code and to discuss how best to proceed. After this, the AWG was re-constituted; the core membership remained but two additional members were appointed, a student from the Students Representative Council and the Head of Registry. The AWG was instructed to complete its task, taking into account all faculty feedback. The AWG then met nine times before May 2001.

Three aspects of the activity in this period were striking.

- (a) Again the majority of the time was devoted to the vexed questions of University-wide grade/class descriptors and aggregation.
- (b) It was decided that an external expert on assessment should be appointed to write the Guide to Good Assessment Practice. An apparently ideal person was approached; she reviewed the material available and decided not to accept the appointment. The reason given was that most of the work had already been done and it would be better for a member of the AWG to complete the task. No member of the AWG was willing to take on the task 'due to pressure of other commitments'.
- (c) The AWG augmented its usual way of working by arranging meetings between three faculty groups (arts-based, science-based and professional-preparation-based) and a sub-group of the AWG. The reason given was that progress was more likely in more-focused groups which took into account 'discipline-based differences in assessment thinking'. The main matters debated at these meetings were (i) whether a common grading scheme was possible for both 'academic' and 'professional-preparation' programmes, (ii) the wording of generic class/level descriptors and (iii) aggregation procedures for both general and honours degrees.

By the end of 2000, the AWG had agreed with the Vice-Principal (Learning and Teaching) a revised form of the Code of Assessment, amended to take account of the relevant parts of the QAA Code, which would be incorporated into the University Calendar to replace the existing Examinations Code. This was endorsed by the appropriate bodies and came into operation for session 2001-2002. Section 2 of this Code says only this: 'A section on standards will be added to the Code subsequently, following approval of that section by Senate'.

Phase 5: September 2001 onwards

The AWG met again in September and December 2001 to try to make progress with the 'Standards' section. The Guide to Good Practice did not appear on the agenda for these meetings.

The post of Vice-Principal (Learning and Teaching) disappeared in June 2002 and the duties were absorbed into the remit of the Clerk of Senate. The Convener of the AWG (an Associate Dean for Education) was reprimanded by his head of department for damaging the departmental rating in the 2001 Research Assessment Exercise by spending too much time on student-related affairs and university committees. The committee structure has been changed; the AWG, if it continues, will in future report to the Academic Regulations Committee rather than to the Committee on Educational Strategy and Resource. The conclusion seems inescapable that what started as an ambitious attempt to generate policy aimed at educational development contracted into the relatively modest production of a regulatory code.

THE NATURE OF THE ASSESSMENT REPORTS

The 'Final Report' of May 1998

In the first ten meetings of the AWG, members talked about producing an assessment policy for the University. As they worked on the report, they returned to the terminology of their remit and thereafter used the terms policy and strategy interchangeably. The May report began by saying: 'The group interpreted its remit as the search for a set of transparent normative principles in respect of assessment which.... command broad agreement across the University and can underpin assessment processes and procedures which are uniform across the University'. The

strategy for assessment would then be the adoption of those principles and the implementation, monitoring and revision of the processes underpinned by them. The principles would later be supplemented by operational guidelines in a document perhaps entitled ‘The Guide to Good Assessment Practice’.

The report identifies four primary and six secondary principles. These deserve careful scrutiny and more extensive comment than is possible here. The following points are significant for the limited purposes set for this chapter; attention is drawn to the frequency of the word ‘should’.

- (a) Assessment should be seen as the exercise of professional, academic judgement about learning, rather than as a collection of methods yielding the evidence on which the judgements are based.
- (b) The primary purpose of assessment from the candidates’ perspective is, quite unequivocally, to contribute to the achievement of learning outcomes. From the institution’s perspective, assessment must uphold and enforce academic standards; this requires the conduct of assessment to be criterion-referenced. Such an orientation depends on transparency in specifying intended learning outcomes.
- (c) Summative assessment should be conducted in relation to clear and meaningful verbal descriptors of the standards of learning outcomes which qualify for the different awards available; one set of descriptors should be generally applicable throughout the University.
- (d) Every course should contain exclusively formative elements in its assessment scheme, in addition to the summative elements required for an award.
- (e) Every assessment scheme should include a variety of modes of assessment.
- (f) All assessment schemes should allow students adequate learning time.

- (g) The assessment process should always involve reasonable arrangements to confirm candidate's individual ownership of the work being assessed.
- (h) Instrumental (key) skills should be assessed separately from component (subject-specific) skills.

Looking back at this report four years on, two things are striking. Firstly, the discourse is clearly educational and privileges one set of concepts over another. The desirable ideas are professional judgement, learning enhancement, criterion-referenced thinking, formative assessment and constructive alignment of assessment methods with intended outcomes. Secondly, the AWG wished broad agreement on the principles and then went on to say these underpinned processes and procedures 'which would be uniform across the University'. Subsequent events showed that it was entirely possible for the whole university to accept the principles, but be strongly opposed to the universal adoption of processes and procedures said to be underpinned by them. For example, the AWG later proposed a grading scheme said to be principle-based; some faculties then fought for their retention of a different and incompatible grading scale which they argued was equally consistent with the principles. When the learning and teaching strategy document was discussed by Senate, the assessment section aroused no direct comment at all, but the document as a whole was seen in two very distinct ways; some said it was inspirational, flexible and acceptable; others said it was dogmatic, prescriptive and unacceptable. The former seemed to focus on the principles and the latter on the procedures said to be underpinned by them. The Principal made the intriguing comment that the document was 'provisional, merely regulatory and non-prescriptive'. He failed to clarify how something could contain non-prescriptive regulations. A possible

clarification, in respect of assessment, is that there are some areas in which it is entirely legitimate for the University as a whole to require that regulations exist and be followed and that there are others where it is reasonable for non-prescriptive guidance to be offered.

The Draft Code of Assessment of March 2000

The AWG had decided by this time that its remit to generate policy (or strategy) could only be discharged through two products, not one. One would be the Code, dealing with the 'mandatory' aspects of assessment; the other would be an 'advisory' guide, stimulating people to think of ways in which their assessment practice might be enhanced.

This code was very much shorter than earlier drafts of 'policy'; much material was excluded on the grounds that 'this is really advice that ought to appear in a guide'. The code contains only eight pages of main text and sixteen pages of Appendices. The language has changed: 'should' has become 'will', 'shall be' or 'must'. For example: 'The assessment scheme will be implemented in accordance with the following requirements....'; 'The head of department shall ensure that...'; 'External examiners must....'; 'The head of registry shall...'

The Code and its Appendices go into considerable detail on what can be called the management and administration of assessment. The responsibilities of a wide variety of people for ensuring that appropriate procedures are scrupulously followed are carefully detailed; much attention is given to the appointment of examiners and the conduct of boards of examiners and to reporting procedures.

What might be deemed the educational aspects of assessment are dealt with in considerably less than one page. The degree of condensation between draft and

final versions was drastic. One member of the AWG (not the present writer) had produced ten pages of carefully argued material on the provision of feedback to students; this was debated at length in the AWG and then agreed. This section was reduced to one sentence: 'The assessment scheme shall detail how candidates will receive feedback to guide their subsequent learning'. Such brevity is not necessarily a bad thing: ten pages give people a great deal to disagree with. More significantly, the AWG had probably been attempting to develop a detailed policy on feedback which could later be centrally imposed, whereas the details of implementation would have been better decided at faculty or departmental level. There is a very important difference here between 'the University policy is that faculties must have a detailed policy on the provision of feedback' and 'the University policy is that faculties must provide feedback in these detailed ways'.

In addition to the short sentence on 'feedback', the Code makes only three brief statements on how student learning should be assessed. These are (i) that performance must be assessed against the stated learning outcomes of the course or module, (ii) that assessment must include an appropriate combination of formative and summative elements and (iii) that assessment must deploy a variety of forms of assessment appropriate to the learning objectives. It could be argued that these very brief statements gain in impact from their brevity and, if taken seriously, could transform assessment practice. The chances of their being taken seriously are reduced when they are immediately followed by many pages on departmental management and central administration.

The QAA Code on Assessment (April 2000)

Obviously this was not a product from the AWG; the AWG was, however, instructed to take it into account. The content and nature of this QAA Code were described in Chapter 2: The National Policy Context.

The majority of the QAA precepts found their counterparts in the AWG code. There is one exception. The QAA required that ‘The institution should ensure that all staff involved in the assessment of students are competent to undertake their roles and responsibilities’. In the earlier drafts of the AWG Code, there were three paragraphs devoted to this matter and both staff selection and professional development implications were addressed. In the condensation of the draft, it was successfully argued that anyone judged competent to be a member of academic staff must be deemed to be competent as an assessor. Concerns about graduate teaching assistants, who carry out an increasing amount of assessment, were dismissed.

The missing products

At the time of writing, two products which should have been delivered have not been. These are Section 2 of the Code (based on the earlier Section 5 – Standards) and the ‘Good Assessment Guide’.

A whole thesis could be devoted to the attempts to produce the ‘Standards’ section. It has been through eight separate versions and has already taken innumerable hours of staff time. What was attempted was the provision of generic grade descriptors for non-honours courses and related generic class descriptors for honours courses, each of which could be related to numerical scores for aggregation formulae. There was a general feeling that throughout the University there should be one common grading scale with rules of equivalence to those different scales that

various sectors would be most reluctant to abandon. Then it became clear that professional preparation faculties and departments, essentially working towards binary decisions between competent and not-yet-competent, felt that the seven (or twenty) point scales favoured by academic subject departments (who believed they could distinguish that number of levels of understanding) were wholly unsuitable. Then the 'harder' academic subjects insisted that any grading scheme must accommodate the percentage scores generated by increasingly used objective and computer-based assessment. These differences could be construed in terms of a tension both between assessing competence and understanding and also between criterion and norm-referenced thinking.

The task of generating seven grade/class descriptors was just the thing to keep academics happily, if not profitably, engaged for hours. Examples culled from other universities gave full rein to people's critical abilities, indeed their negativity, and submerged any obligation to be creative. Although members of WGA remained commendably amiable, two distinct sub-groups emerged. The first maintained that the task was impossible ('there is no point in forcing a marriage between approaches that are conceptually incompatible') and that the group had to begin again on some rationally defensible task – even if that meant being much more independent-minded and radical. The second argued that although the task might be theoretically ill-conceived, some practical resolution had to be reached; the effort was worthwhile if it resulted in a 'least bad' solution, one which was workable if not ideal.

The second non-product was the Guide. This was envisaged as a carefully developed and persuasively written document which would encourage people in all parts of the University to evaluate their current assessment practice in well-informed

ways and to improve that practice in ways acceptable to them and appropriate to subject-based differences. The document would be made vivid and readable through case-studies of successful innovation; it would contain a glossary making assessment 'jargon' accessible; there would be a select annotated guide to further reading and research; sources of support would be identified. The Guide still exists only in skeletal form; no-one has been found who is willing to develop it.

THE ASSESSMENT WORKING GROUP: CRITICAL COMMENT

Between January 1998 and January 2002, the AWG had forty routine meetings; in addition there were special consultative meetings, either large meetings with faculty representatives or smaller ones with sub-groupings of faculties. Any attempt to cost this whole exercise would have to take into account not only the time of academic and other staff at meetings, but also the time put in by both categories **between** meetings. The cost of paper and photo-copying alone must have been huge, notwithstanding all the attempts to communicate and distribute documents by electronic means; the present writer, for example, has accumulated a pile of paper (the 'dead-tree documents') seventeen inches high. Examination of the products of the AWG that were actually used, and considering what was **not** produced, only two summary judgements seem possible: either the AWG signally failed to do what was required of it or, less harshly, it failed to give value for money. There are various explanations; five deserve some scrutiny.

(a) *The members of the group were not up to the task.*

The writer was told on one occasion, 'You're a bunch of typical academics – all brains and no common-sense' and on another occasion, 'As we say in Partick,

you lot couldnae run a ¹minodge'. The intellectual calibre of members (allowing one exception) was impressively high; all had substantial experience of assessment practice; the Convener was distinguished by the subtlety of his intelligence, his sound grasp of the big issues and his knowledge of the internal workings of the University. Criticism based on lack of ability would be hard to sustain. Rather more valid would be the criticism that the group regularly failed to realise when the pleasures of debate should give way to the need for decision-making. The Convener often said things like, 'That was a most enjoyable discussion, I suppose we'd better get back to the task in hand'.

(b) *Members were unable to see the bigger picture.*

In phase 1, members had no difficulty in handling big ideas; there was no great difficulty in explicating normative principles to underpin all assessment practice. However, when it came to specifying processes and procedures, the criticism of inability to see the bigger picture acquires some force. In the notes of meetings were recorded frequent statements like these. 'I'll outline what we currently do in maths and computing; anything that fundamentally threatens current practice will not be acceptable to my colleagues.' 'This is what we have done for years in life sciences.' Energetic defences of the status quo and of sectional interest were commonplace. It can also be argued that the interests represented in the AWG were biased towards the 'hard' sectors within the four-quadrant model and members were unlikely to deal acceptably with arts and social sciences. The researcher sought reasons for the composition of the AWG; he was told that people had been chosen on their individual merits. They were people well-known to be keenly interested in,

¹ A minodge is a small local savings club; the term comes from the French 'ménage' (house-keeping).

and knowledgeable about, assessment rather than those who would act as representatives of different subject areas. It could of course be the case that AWG members were perfectly capable of seeing a bigger picture, but were very aware of the wide range of possible difficulties in projecting such a picture; there is, however, rather a fine line between cautious awareness of potential objections and a failure of nerve to propose radical action. Re-reading notes of meetings, this writer is aware of his own failure to push hard enough for what he believed to be sound; the fact that he wished to continue in his research role is his excuse.

(c) *The bigger picture was too monochromatic.*

The AWG spent much time trying to formulate assessment processes and procedures 'which would be uniform across the University', before reaching agreement that some processes/procedures should be uniform and others should not. With the benefit of hindsight, the slowness of this realisation seems lamentable, but it stemmed from a genuine desire to think of the University as one, great institution with its own distinctive character and some unity of purpose. This perception, it must be said, was engendered and reinforced by messages coming from both Senate and central management. At an early stage, one authority figure drew a revealing comparison: 'Assessment practice is like the University's architecture – a mess. Things have just been added on with no thought for the overall picture. You must do better than this...'

(d) *The initial remit was either ill-conceived or cynical.*

The AWG developed and maintained a feeling of group solidarity which persisted through internal disagreements; in part this was due to the inter-personal

skills of the Convener, but in part also it was a defensive response to a perception that it was being 'mucked-about' by central management. The argument was that the University was required to do **something** by emerging national policies so, without giving adequate thought to the achievability of the remit, it set up a committee, the AWG, 'just so that they can say they have responded positively'. There then developed a feeling that central management was very ready to commandeer any AWG product that it liked and present it as its own. When the AWG appeared to be about to recommend something **not** to its taste, support was withdrawn. One member observed (in April 2001): 'They gave us an impossible task, kept us in the dark and threw shit at us when we tried to grow the wrong kind of mushrooms'.

There is no convincing evidence that the AWG was persecuted, but there is evidence of occasional paranoia. There is no evidence of cynical conspiracy amongst the central authorities of the University; there is, however, evidence of a lack of forethought by those setting up the AWG, both in respect of remit and membership. (Interview evidence revealed that, for example, the Vice-Principal's recollection of the remit bore very little resemblance to the actual remit.) If someone had a clear vision about what should be achieved, that vision was never conveyed to the AWG. The amount of time spent by the group agonising over what it was supposed to be doing was deeply troubling.

(e) The implied model for stimulating change was flawed.

Some people at the centre of the University, probably in response to external policy requirements, decided that certain changes in assessment were necessary. They set up a small group, if not of the great and good then of senior and influential

people, and charged the group with devising 'policy'. There was an expectation that a helpful, consistent and authoritative policy would emerge; there would then be some routine consultation, and then after minor revisions, the policy would be implemented. The model implied was of central determination and peripheral implementation i.e. a top-down centre-periphery model. One external observer commented that this was only to be expected of an ancient Scottish university known for its very traditional, authoritarian character. Another justified the approach by saying there was so much inertia in the system that any significant change could only come by the application of force from the centre (a 'boot up the arse' was the actual phrase).

Whether or not it should have been seen from the start that the change strategy was likely to fail, there were three lessons that AWG learned over four years of low productivity. Firstly, in a diverse and complex institution of higher education, a crude attempt at vertical policy-making (see Chapter 2) without any attention to the horizontal processes of policy-activity was ill-considered. Central management appeared to have given no thought at all as to how change might best be managed. Secondly, whether people espoused a top-down approach or a bottom-up one, the crucial influence of the middle had to be recognised; in the middle ground between central initiatives for change and the actual practice of individual academics, the contribution of faculties/ departments was highly significant. Thirdly, assessment embraces so much that an ill-focused attempt to bring about widespread change ended up pruning out concern for the enhancement of student learning and focusing almost exclusively on regulatory, procedural, administrative and bureaucratic concerns. Similar lessons have been learned elsewhere.

First, top-down assessment efforts are not all that fruitful in terms of creating a positive culture of assessment. A bottom-up approach may not be possible without great expenditure of educational effort. Some kind of balanced strategy may be ideal. Second, when assessment is perceived as being (or actually is) driven by external forces (accreditation, government accountability or whatever) the institution's eyes tend to focus higher up the organisational chart than where the students' learning occurs.

(Borland, 2002)

THE PERSPECTIVES OF TWO KEY INFORMANTS

As described in the previous chapter, the Vice-Principal (VP) responsible for setting-up the AWG and its Convener (Con) were interviewed. These semi-structured interviews were tape-recorded and based on an outline schedule strongly influenced by research questions 1 and 2. The main areas covered were (a) the meanings attached to key terms – policy, strategy, code and guide, (b) what aspects of policy should be centrally prescribed and (c) ways of encouraging assessment change.

Meanings attached to key terms

In the early days of the AWG the terms policy and strategy were used interchangeably. This was not quite as the interviewees saw it retrospectively. For one, the terms were synonymous and for the other there was a distinction: policy was concerned with purposes, principles and priorities whereas strategy was concerned with how policy should be implemented. Both views are in harmony with standard dictionary definitions; one of these defines strategy simply as policy, another differentiates between principles of action (policy) and a plan of action (strategy). Both interviewees saw policy as something generated and then implemented i.e. policy was construed as policy-making rather than policy-activity.

One saw policy as being served by two things: a code (a set of mandatory regulations compliance with which was essential) and a guide (a set of advisory recommendations about good practice). The other conceived of policy as requiring three components: a set of principles, a code of (mainly) requirements and a compendium of advice. The code should not be seen as exclusively requirements; it should also explicate for the organisation its philosophy and common purpose.

In a large, complex institution, there is some uniformity of purpose, of common purpose. And the job of the code is to express that common purpose and this is not merely expressed in doing the same things through a common set of regulations. The code is rather more than just a set of requirements. It is an expression of a philosophy and a commitment. Part of our problem... is that I am not altogether sure that the faculties have begun to think in the right kinds of terms to discuss this issue. I think they have assumed that a code is a common set of regulations and thus have seen it as endangering their existing procedures; rather than that what we are trying to work towards is something which expresses a commonality of purpose.

[Con]

Although only one insisted that assessment regulations be justified by subscription to a common educational purpose, this did not mean that the other saw the code as merely a **collection** of regulations. He emphasised that the regulations would have an internal coherence and would remove unacceptable inconsistencies in practice. It is interesting that many standard definitions of 'code' stress the compulsory status of the components (laws, rules, commands, regulations) and a few stress that they form an organised body of such components 'so arranged as to remove inconsistency and overlapping'.

What, in assessment, should be centrally prescribed?

Although the two informants appeared to respond in different ways, in essential substance their answers were similar. Three common themes emerged.

(i) There were things that should be centrally prescribed and that the University had an obligation to prescribe. These were agreed to include: the acceptable form of reporting results to the Registry; the eligibility of staff to act as assessors; procedures for the appointment of external examiners; the membership and conduct of boards of examiners; the provision for special needs cases; the handling of disciplinary cases in which assessment regulations were contravened.

(ii) In some matters, the University could require faculties to articulate policy and in addition could prescribe that whatever the policy was in detail it should embody a certain principle. Two examples were given. Firstly, University policy now required faculties to base their assessment schemes on the principle of constructive alignment of assessment methods and intended learning outcomes; the details of how this would be achieved were matters that could only be decided at faculty level. Secondly, faculties must provide students with constructive feedback but how that was to be done was for them to decide.

(iii) In some matters, central prescription would be entirely inappropriate. What should be assessed in a certain subject area, and the methods used to assess it, were seen by both interviewees as matters completely outwith the competence of any central authority, or centrally derived policy, to prescribe.

These distinctions may seem clear-cut, but there were difficulties in particular cases. One obvious matter that had exercised the AWG and both interviewees was the extent to which one common grading scale could be required.

This could be seen as a matter of reporting to the central registry (in which case the University could define a uniform policy); it could also be seen as deciding on grading appropriate to the particular educational purposes of different subject areas (in which case the University could not require uniformity). In later chapters, it emerges that different subject areas made a distinction between bureaucratic and educational aspects of assessment; however in some matters the two are intertwined. Another difficulty was that staff in departments were prone to confuse what the policy actually was with what they perceived it as being and indeed to believe that a formal policy existed when it didn't. (There was a memorable occasion when the Vice-Principal had said, perhaps jokingly, 'In the absence of any written policy, what I say is policy'. None of his audience had laughed.) It was not always easy to distinguish between a central prescription and a central 'push'; central management had a variety of ways of sensitising people to what should be given priority. The interviewees had much to say about the appropriateness of central prescription in assessment. The themes above are illustrated below.

Would it be reasonable for the University to have a prescriptive policy on this? It all depends what you mean by the University. If it includes faculties – yes. It would be quite reasonable to have different prescriptions in different faculties. We do that. The rules in medicine are different from the arts. [VP]

There is also, I suppose, below the level of a rule, a sort of push, for want of a better word, towards using a wider variety of assessment methods. It wasn't written down but there was a sort of feeling that this was the kind of thing people were expected to do and part of that was frightening people with quality assessors. You know the kind of thing... 'the quality assessors are going to be surprised if you don't...' This is at the level of persuasion rather than coercion. [VP]

Part of the common philosophy of concern for standards has at some time to confront the issue of a common language to talk about standards and I don't

have any difficulty about prescription in that area. However, I certainly uphold the view that standards have to be applicable to the different subject areas concerned. [Con]

You can prescribe that assessment judgements be aligned with learning outcomes – but that does not endanger academic freedom. The individual department, the academics, have complete autonomy to determine what the learning outcomes of a specific course are to be and to elect to apply an assessment instrument that is appropriate for the assessment of the stated objectives. The code is couched in sufficiently general terms not to endanger autonomy. [Con]

Ways of encouraging assessment change

Perhaps the purpose of the AWG was simple i.e. to generate policy which would change assessment practice in desirable ways. This was not quite as the interviewees saw it.

One said that ‘Hopes for the AWG were focused primarily on regulations. What it probably wasn’t meant to do was to look more generally at assessment and make it more useful.’ Existing regulations were ‘in a terrible mess’ in that there was internal inconsistency and many matters that were just not covered; the language was often inappropriate, concentrating narrowly on examinations rather on the wider idea of assessment. This was not then policy-making to bring about change, but rather a housekeeping exercise to tidy up the regulatory code. In one respect only, the AWG was expected to go further; this was to achieve grading that was more in line with statements of intended learning outcomes. On probing whether this was not in fact policy intended to improve practice, the reply was as follows.

The working group was not set up to initiate change. It was about bringing our rules and guidelines up to cope with the situation we had already arrived at. I suppose where there is a bit of a push there, it is in areas which lag somewhat behind, but I don’t think there was any pushing of practice beyond the broad average of what was already taking place. [VP]

This perception of purpose was not shared by the Convener. In his view the AWG was meant to achieve three things. Firstly (agreeing with VP) a code should be produced; this would be more explicit and coherent and less easy to contravene. Secondly, agreement was to be reached on a set of principles which, 'if properly understood', would produce a general advance in assessment practice. Thirdly, there would be a guide that would be innovating in two senses: getting people to think more carefully about assessment, but also 'innovating in the sense of offering suggestions as to different ways of approaching the operation of assessment'. The difference between these two perspectives was disconcerting. It was as if one person put a minimalist interpretation on the remit, the easier to claim that the AWG had achieved some success; the other clung to a broad interpretation and was prepared to live with the judgement that 'his' AWG had failed. Less time would have been wasted either if the remit had been more carefully composed at the outset or if it had been re-formulated after phase 1.

In a chance conversation with another senior manager, the researcher asked if it was his understanding that the AWG had been set up to generate an assessment policy. The answer was, 'Not so much to generate policy as to provide a lead in improving assessment practice'. How then is leadership in assessment to be construed? Yorke (2001) has argued that 'the leader needs to go beyond ensuring routine compliance with external and internal regulatory frameworks to identify ways in which assessment practice might be developed'. This 'going beyond' the AWG did not achieve.

The two interviewees may have disagreed on whether the AWG was intended to encourage any general innovation and improvement in assessment. They were of one mind, however, in saying that it was necessary for a new policy more

clearly to proscribe what was unacceptable. Examples of unacceptable practice were given: one department had an inadequate policy on re-submission of coursework which had resulted in students re-sitting examinations they had already passed; another did not allow external examiners to see examination papers in advance; in another, meetings of a 'board' of examiners had taken place at which no internal examiners were present. Such practices were impermissible; any code must prevent them recurring.

The interviewees had different views on the role of external policy bodies. One emphasised (a) that the fact that very similar changes in assessment had happened elsewhere was evidence of direction from external policy and (b) that the University had set up an assessment policy group because the QAA, post-Dearing, made it essential and any institutional policy would have to be clearly in line with a QAA Code. The other argued that similarities between a Glasgow code and the QAA Code were coincidental rather than a matter of direction; he believed the University had retained its autonomy.

The QAA Code of Practice does include a section on assessment and the WGA is familiar with the content of that and has gone out of its way to ensure that it accommodates the precepts of the QAA Code. But I would argue that it has done so coincidentally, although explicitly, rather than the Glasgow code being manipulated by the QAA as it were by a force acting at a distance. [Con]

What then did these two people see as productive ways of bringing about a general enhancement of assessment practice?

(a) A Code would have the limited role of bringing the below average up to an acceptable standard of procedures (presumably also pushing up the average in the process!)

(b) One saw policy as being influential in that it required acceptance of underlying principles and attention to a guide to improving practice.

(c) One stressed that advance would come about, not through policy, but when the climate in departments and faculties fostered the enthusiasm of individual lecturers. There was a huge task here for educational developers (like, for example the Teaching and Learning Service) probably not in running courses but working hand-in-hand with staff in departments on an outreach basis.

Possibly the most striking thing of all came from the Vice-Principal. Debate on the big issues in assessment had previously been expected to occur within the Teaching and Learning Committees of departments and within the Education Committee of Senate. As a result of the policy-activity stimulated by the AWG he had decided that Teaching, Learning and Assessment Committees must be established at faculty level. This should help ensure that sufficient importance was attached to the development of assessment policies and practice; the faculty committees would require departmental committees to be more active in this regard than they had previously been.

DID THE UNIVERSITY GENERATE NEW ASSESSMENT POLICY BETWEEN 1998 AND 2002?

If the regulatory Code can be construed as a policy, then it did. However, that Code remained incomplete; discussion of the crucial section on standards, class/grade descriptors and aggregation procedures still continued in June 2002. If

policy is construed more widely, then it did not. At the time of writing this thesis the Guide to Good Assessment Practice has not progressed. Agreement has been secured to a set of underpinning principles; these are sufficiently vague to allow widely differing interpretations.

On the other hand if policy is conceptualised as process rather than product, then there is no doubt that there was generated an unprecedented amount of policy-activity centred on assessment. A large number of people were sensitised to assessment issues; assessment practice rose higher on the agenda than ever in the past. The importance of assessment was recognised in the setting-up of a new policy-determining structure of teaching/learning/assessment committees. What effect these developments will have on future policy-making is unknown. Perhaps the most optimistic shift has been away from dependence on two unproductive change strategies: a simplistic, authoritarian top-down approach and an ineffective, bottom-up, 'leave it to innovative individuals' approach.

DID THE POLICY AIM TO CHANGE ASSESSMENT PRACTICE?

The emerging policy aimed to make unacceptable practice less likely; it is a matter of some disagreement whether it intended to be about some general enhancement of practice. If an admittedly naïve distinction between administrative and educational aspects of assessment is accepted, then emerging policy overtly aimed to change the former. The position with respect to educational aspects is less clear-cut.

Emerging policy prioritised a few educational principles. In the early stages of policy generation there was a belief that adoption of these principles would lead to 'processes and procedures which are uniform across the University'. Over time

there was a discernible shift in AWG thinking. Simply put, it moved from ‘policy will require you to implement A, B and C’, towards ‘policy will require you to address this principle in ways appropriate to your subject-based situation’; the distinction between policy as a plan of action and policy as principles of action became clearer. With the increased emphasis on decision-making at faculty level, there was an increased insistence that direction from the centre would not erode subject-based, professional autonomy. As the Convener put it, ‘Where there is specific need there should be specific provision. There should be uniformity in the sense of unity of purpose, but not identity of procedures.’

The rhetoric of the policy-makers in the latter stages of the AWG was firmly based on the premise that there should be no more central direction in assessment than was necessary for legitimate University purposes. In evidence to be discussed later, academics in different subject areas revealed their fears that policy direction from the centre would increasingly ‘interfere’ with their assessment practice. Were they right to fear that the reality would be different from the rhetoric? The answer provided by the Vice-Principal provides little comfort.

I think there is an unjustified fear of central control across the University. That wasn't there thirty years ago. Why that fear? I suppose because there is a degree of central control that wasn't there in the past... There are more rules now and who knows what it will be like in ten years time? [VP]

CODA

This chapter has illustrated that the subtleties of the internal processes of the AWG could never be imagined from the sterility of its formal products. What Colebatch (1998) has called the ‘profane knowledge’ of the participants could not be stated in the ‘sacred discourse’ of their public pronouncements. At the time of

interview, the AWG and its Convener appeared to believe that they had been engaged in authorised decision-making which would have a significant impact on assessment practice. In other words, we have here an essential policy myth, defined by Yanow (1996) as ‘a narrative created and believed by a group of people which diverts attention from a puzzling part of their reality’.

Two things were clearly demonstrated: in policy development agora and oikes are as important as ecclesia; policy-making is graduated and problematic rather than definitive and absolute.

CHAPTER 5 CHANGING ASSESSMENT PRACTICE: CHEMISTRY

INTRODUCTORY NOTE

Readers are encouraged to study the full chemistry report, Appendix IV. The content of this chapter is extracted from that report and supplemented by material from the data file. The treatment begins with a note on chemistry as a subject area and then gives an account of how interviewees located the subject within the framework described earlier. This addresses research question 8. The bulk of the chapter is then structured around research questions 4, 5 and 6. After this, emerging issues important to chemistry staff and issues imported by the researcher are described; this contributes towards answering research question 7. The same structure, *mutatis mutandis*, is adopted in the next three chapters.

WAS CHEMISTRY SEEN AS ONE SUBJECT AREA?

All those interviewed agreed it made sense to treat chemistry as one broad subject area; more than half said that although they saw it this way, some of their colleagues did not.

HOW DID PARTICIPANTS LOCATE CHEMISTRY WITHIN THE 'FOUR QUADRANTS' FRAMEWORK?

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 was genuinely close correspondence between the rhetoric of course purposes and the reality of assessment procedures. Confirmation of this would require not only an analysis of what was sought through assessment instruments, but also some scrutiny of what was rewarded in student responses. This goes beyond the present study.

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). Honours chemistry was thus thought to put a significant emphasis on assessing understanding of a defined set of key concepts and to put slightly more emphasis on the 'pure' rather than the 'applied'. Chemistry was seen as 'hard and pure', but in a less decisive way than hypothesised.

WHAT SIGNIFICANT CHANGES HAD OCCURRED? WHY?

How students were assessed and what participants saw as pleasing are described in detail in Appendix IV.

Interviewees were asked to identify what they saw as significant changes in how students were assessed. People were not restricted to some arbitrary period of time, but encouraged to reflect on the whole period over which their experience extended. They were then asked why they thought change had occurred. It was anticipated that the latter would be a difficult question to answer and so it proved. Change 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.

Level 1

Interviewees identified four major changes they saw as significant. Firstly, continuous summative assessment had come to play a greater role; secondly, the introduction of short diagnostic tests had motivated students to work more

consistently; thirdly, continuous assessment served a useful formative function in that struggling students were identified earlier and remedial support was more effectively targeted; fourthly, the assessment of practical work now received serious attention.

Taken together these four changes say much about a general shift in thinking about the big purposes of assessment; there was increased recognition that assessment serves not only a certification/accountability function but also an educational/learning enhancement function. Neither the extent of this shift nor the ease with which it had been achieved should be exaggerated. For example, changes in practical work assessment were thought significant, but staff found some difficulty in describing them. People said its weighting had been increased, but could not recall what it had been or how great the increase was. The current modest weighting to practical work (10%) was thought about right; although students might press for an increase in weighting, staff would oppose this. There was a view that assessment of practical work was very subjective: tutor-demonstrators tended to be inconsistent in the application of assessment criteria. Others believed that the considerable time invested in clarifying the purposes of practical work had been of real benefit in that it had resulted in not only the explication of assessment criteria but the introduction of pre-laboratory sessions. This seems a good example of an assessment method favoured on grounds of its apparent validity and authenticity, but at the same time distrusted for its low reliability; it also reveals the tension between criterion-referenced thinking (testing for mastery of skills) and norm-referenced thinking (testing to spread people out).

Honours

The general view was that assessment at honours level within BSc chemistry courses had not changed significantly: 'We haven't tinkered with it much, because it seems to work well'. Infrequent mentions were made of three changes considered of some significance: the increased weighting given to the Honours research project; the introduction of carry-over assessment from Junior to Senior Honours; the 'tightening-up' of the assessment of laboratory work. Two single mentions occurred of changes each highlighting problems of aggregation: the separation of second-class honours into two divisions and the introduction of designated degrees. In the former case, aggregation was said to have become more formulaic due to an increased dependence on 'the manipulation of numbers with spurious validity in the interests of transparency'. In the latter case aggregation had to be carried out 'in a bureaucratic way which paid scant regard to educational coherence'. Both interviewees regretted the reduction of scope afforded them, and boards of examiners, for 'professional judgement' in deciding final awards.

Within the new MSci courses the big assessment change had been a diversification in the assessment methods used; assessment procedures now included essays, poster presentations, assessment of language competence and placement reports. The last of these involved, for the first time in chemistry, collaboration with placement supervisors with a defined assessment role.

The reasons for change

Evidence came from two distinct parts of the interviews: people were first asked 'Why, in your view, did change occur?' and later they were presented with a

list of factors influencing assessment and asked how important each had been in the changes described by them earlier.

(a) Interviewers' views

It was hoped 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 (given the nature of causality in human affairs) or the interviewing technique was inadequate to the task. What happened in most 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 always 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. Thirdly, there were changes in assessment practice which were part 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 was considered and it was likely that significant changes would be introduced. The creation of the four-year MSci course was an example in this third category.

The reasons given for these changes fell under three headings: learning enhancement, conformity to fashion and expediency. Learning enhancement in

some form or another was mentioned most frequently, but it was noticeable that most people tended to temper one 'worthy' student-centred factor with some other factor they themselves described as 'less educationally respectable'.

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.

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

The reasons categorised as expedient were different from 'mere' fashion. They included things like the following.

We did that just to get better pass-rates.

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

Apart 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.

(b) Interviewees' reactions to the views of others

Presented with the list of fifteen factors influencing past changes in assessment, interviewees reacted as summarised over.

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

Important points made when amplifying responses follow.

- (i) Although general concern for the quality of student learning was seen as the most important causal factor, the main theme in the comments was 'students working harder and learning more'. Interviewees stressed the **quantity** of learning rather than its quality; there was little overt recognition of the potential of appropriate assessment to encourage higher cognitive abilities or deep rather than surface learning.
- (ii) The Royal Society of Chemistry was thought of great importance. However, it had not acted as a stimulus to change; rather the department only initiated change that would be acceptable to this body.
- (iii) Whenever a new course structure had been introduced, it had been a policy-required part of the course-planning process to review assessment.
- (iv) Attention to transferable skills and their assessment was sometimes said to have been a sincere response to an internal departmental concern about the

employability of their graduates and sometimes to have been a cynical bit of window-dressing to meet an external 'push'.

(v) Change in assessment practice sought by the department needed the approval of external examiners; the department had always been able 'to find good reasons to resist any recommendations for assessment change' made by externals.

(vi) Increased diversity within the student population had no direct impact on student assessment. The indirect impact was significant. The response to increased diversity in student abilities and needs had been the creation of new courses; in creating these more thought had been given to assessment.

(vii) External policy requirements, whether national, institutional or faculty, were seen as having had little impact. This was also the opinion on assessment trends elsewhere, assessment technology, student pressure, student numbers, the influence of enthusiasts within the department and of educational/assessment theory. The last of these prompted more revealing comment than the others: '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'.

The eight people shared a common subject and departmental culture. Could they then be described as an academic 'tribe'? There was some commonality in views and attitudes, but the eight clearly showed their individuality and revealed their differences. Perhaps, however, they belonged to different sub-groups within the tribe? It was decided to re-interrogate the data with two 'hunch-hypotheses' in mind: one, that the views of older, more experienced staff were distinct from those of younger staff and, two, that the three women were different in their assessment thinking from the five men. There was no convincing evidence to support these

hunches; indeed the two people most similar were the most senior man and the most junior woman.

Things work well enough but they could be a lot better. There are many assessment issues that are in need of attention, but we don't seem to have the time or the appropriate forum to deal with them. And I feel I have to say that some of my colleagues are not always open to persuasion on assessment matters.

It takes a lot more time to improve assessment than just to do it. And time is something that we don't seem to have. I want to talk with colleagues about assessment a lot more; there is some chat in the coffee-room, but there is no real debate that might result in things being changed. You get the feeling that the old-hands don't really see any need for change.

WHAT FURTHER CHANGES (a) DID STAFF WISH TO SEE AND (b) THINK MIGHT BE REQUIRED OF THEM?

Changes wished: Level 1

Two of the eight did not wish to see any further change because 'assessment is currently very satisfactory'. There were four changes desired, each by at least three people (and here the most strongly held views **did** come from the least experienced members of staff). Firstly, core knowledge ('what students **ought** to know') should be more explicitly defined and assessed; this was wholly consistent with the location of chemistry in the four-quadrants model. Secondly, there should be departmental learning and decision-making about what could appropriately be assessed by objective and computer-based testing. Thirdly, laboratory work should be more reliably assessed. Fourthly, there should be more effort to explicate assessment criteria and to achieve common understanding of them.

Running through many contributions was a plea for more time for assessment: for marking, for supporting students diagnosed as weak and to allow more departmental discussion of how assessment should change.

Changes wished: Honours

Four did not wish to see any change, again because ‘we do assessment well’. Pleas for defining and testing core content were repeated and again there was a wish for more rigorous assessment of practical skills ‘because of their importance to industry’. One interviewee identified six ideas for change she felt strongly about; such ideas were sometimes ventilated in the coffee-room, but ‘I have no idea what gets discussed in the Learning and Teaching Committee’.

Changes that might be required

Interviewees were asked what future changes in assessment they thought might be required of them, whether they would welcome them or not. Staff did not speak at length; they saw little to fear and 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. In a clear difference to their reflections about the past, 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, one was viewed in a definitely positive light: the University might require the department to do more to detect and 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 coming from 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 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 methods – agreement on what's ideal is pretty unlikely.

There was one thread running through most predictions: external policy requirements always seemed to be in the direction of intensifying the '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.

The expectation that 'policy' would become much more significant in the future was confirmed in responses made during the structured part of the interviews.

Factors expected to be more influential in future

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

WAS POLICY PERCEIVED AS INFLUENTIAL IN BRINGING ABOUT ASSESSMENT CHANGE?

From what has already been written, the obvious answer would seem to be 'not much in the past, expected to be more in the future'. This needs closer scrutiny.

Interviewees said that the major influence bringing about change was the general view of chemistry lecturers that a change in assessment methods would result in more work and more learning by students. Internally generated proposals for change had to be in line with the policies of the subject- professional body and the Science Faculty. Major changes in assessment had occurred when there was a radical change in course structure. However, there were University requirements as to how course planning must proceed and these had been strongly influenced recently by policy from external sources. Whereas in the past it had been normal for courses to be planned in terms of content coverage with assessment 'bolted-on' at the end, this was no longer the case; assessment had to be considered as an integral aspect of course-planning. Where assessment had changed to make more provision for the assessment of transferable skills this was attributed to 'something in the zeitgeist' or to a push from 'people up the hill'. Those who amplified this thought the influence stemmed from national policy statements on higher education which were then 'handed-down' by central management.

Staff within this subject did not say 'there was an external policy which required specific changes in assessment practice'. Rather they said there was some 'push' from outside which encouraged change and influenced the nature of changes contemplated. They were not much concerned where the policy originated; they did, however, distinguish between external bodies which were subject-based and those which were not.

The future would be different. Put crudely, interviewees expected to be told more often how their students should be assessed. This expectation appeared in three forms. Firstly, central direction (from senior management in the University) would extend into new aspects of assessment. University policy had always dictated what had to be done in the 'administrative and bureaucratic' aspects of assessment; this direction would extend to assessment methods, to a common grading scheme and to uniform aggregation procedures. Secondly, interviewees predicted more 'interference' from national policy-generating bodies, vaguely 'the government' and more specifically agencies like QAA. Opposition to national bodies was strong when they were perceived as 'knowing nothing about chemistry'; unwarranted interference from QAA might well be legitimate intervention by RSC. Thirdly, staff expected policy to become more directive. The department took pride in the way it had resisted policy pressure for modularisation of courses. This had been an externally-derived policy, with clear impact on assessment, which had allowed some measure of resistance and re-definition. Future policy, it was feared, would 'give us less room for manoeuvre'.

EMERGING ISSUES

The Burden of Assessment

Staff talked a lot about the overall assessment workload, but they did **not** complain about it. The characteristic view was:

Assessment is time-consuming but not too time-consuming. I myself find marking wearisome and dispiriting, but it is just something that has to be done – a necessary evil.

This lack of complaint might suggest that the assessment burden was not really an issue at all. However, (a) it was predicted that assessment workload would

soon become a serious problem (due to increased numbers of students and the increasing proportion who were 'critical and litigious customers', to an increased diversity of assessment methods deployed summatively and to a squeezing of teaching/assessment time by the research commitments of academics); (b) a few said that although they did not complain, their colleagues increasingly did; (c) 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.

Plagiarism

The majority were concerned about plagiarism. It was recognised that there had always been a minority of students guilty of some dishonesty in relation to assessment and that the present situation was not a crisis. Nevertheless, there was a clear view that plagiarism was on the increase and this was for two reasons. Firstly, there was increased temptation for students to take short-cuts when more and more of them were undertaking part-time work to avoid financial problems; secondly, students were inevitably affected by the decline in standards of honesty in wider contemporary society.

'Newer' forms of assessment provided more opportunities for student dishonesty than traditional examinations. Assessors could be 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 to be encouraged, but some students benefited inappropriately by presenting as their own the work of 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 was more troublesome in word-processed products in which other people's work could be copied and pasted. Interviewees spoke of the need for vigilance in spotting work which had been purchased or simply down-loaded from the Internet.

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

Objective Testing

For complex historical reasons and because they felt that the nature of their subject required constructed-response rather than selected-response assessment, chemistry staff were opposed to multiple-choice tests. Nevertheless, they felt that the potential of objective testing had not been adequately explored. In their responses, interviewees often revealed a surprising lack of knowledge and understanding of what objective testing meant and the range of techniques (and technology) available; some recognised this deficiency. ('You people should put on a course for us.') There was a recognition that tackling this issue would take time and time was not easily found. This was the most obvious topic in this subject area

where there was strong evidence that assessors lacked something needed 'to do good work' in assessment.

Examination Stress

This was an issue of passionate interest to a few, rather than one raised by the majority. It was recognised that terminal 'big-bang' examinations could cause students considerable distress; there were diametrically opposed arguments on what to do about it. One person argued that examinations were 'a very artificial cause of stress, unlike anything met with in real-life' and thus there must be more diverse forms of continuous, summative assessment. A second argued that coping with this type of stress was a necessary preparation for the demands of adult-life and thus staff should not acquiesce in that foolish tender-mindedness which reinforced a decline in the values of endurance and responsibility.

On re-visiting these data, the researcher was struck by three things. One, these two people should have been discussing the issue together. Two, discussion would be more profitable with knowledge of relevant assessment research. Three, a factor in judging assessment should be its effect on later learning. This is part of the argument for 'sustainable assessment'. (Page 228)

IMPORTED ISSUES

Dominant purposes of assessment

Two people concentrated on the accountability purpose and one on the certification function. Five acknowledged the certification purpose, but chose to give greater emphasis to learning enhancement.

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

Constructive Alignment

Only one interviewee (with experience of committee work outwith the department) used the discourse of constructive alignment. One talked exclusively in terms of conventional content alignment. The remainder focused on epistemological alignment i.e. assessment was carried out in ways considered true to the nature of the subject. The weakness of this position as a defence of traditional examinations, still the norm in a wide range of academic subjects within the University, was not recognised.

Communicating about Assessment

There was an impressive amount of collaboration in the construction of assessment instruments; increasingly this did not involve face-to-face discussion, but rather the circulation of drafts for e-mail comments. There was much discussion in boards of examiners but this focused on the aggregation of marks (seen as increasingly formulaic) and on decisions about border-line candidates; these boards were not the places to discuss the 'big issues' in assessment.

There was evidence of shared understanding about standards and assessment criteria amongst the more experienced members of staff; this had developed through interaction over many years. Less-experienced staff felt more discussion of the meanings to be attached to criteria was necessary. Some people mentioned there was a little (but not much) discussion of assessment within the Teaching and Learning Committee; others mentioned conversations and arguments in coffee-room

and corridors. The two groups were somewhat distinct and each seemed unaware of what assessment issues mattered to the other.

There was no evidence that assessment procedures and issues were ever discussed with students. Self and peer assessment were never mentioned.

Re-visiting the data after an interval the researcher was convinced that for these eight interviewees in chemistry, assessment was something that was **done** but not much talked about.

Criterion-Referenced Assessment

The heading is shorthand; the issue is the extent to which assessment practice was described as criterion-referenced rather than norm-referenced. It is doubtful whether any actual assessment scheme is wholly one or the other; in practice pure forms do not exist. It is worth noting that in those higher education institutions in Australia in which official policy now **forbids** any form of norm-referenced assessment, there are 'powerful remnants of norm-referencing still lying around'. (Boud, 2000)

Within this group of interviewees three people, disliking the complexities and unresolved dilemmas of hybrid thinking about assessment, were clinging to earlier simpler ways of thinking, and using the discourse of norm-referencing exclusively. The majority recognised the transitional nature of their current arrangements and were struggling to find the right language to describe their attitudes to aspects of them. There were clear differences as to what was thought desirable.

Almost all of the students were getting nine out of ten [in tests of practical work]. We weren't happy about that, so we made the criteria more rigorous to spread students out.

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

Constrained Autonomy

Autonomy in assessment is problematic. When a member of academic staff is actually assessing a piece of student work, s/he is essentially autonomous; there is freedom to act independently of others. The responsibility involved in taking discretionary decisions in difficult areas is recognised as a professional obligation. That there is scope for independent professional judgement does not mean that the assessor is free to indulge in whimsical, arbitrary and biased ways; there are obvious constraints imposed by assessment criteria, by rationality and by fairness. Individual autonomy is properly constrained by such things.

The idea of academic staff in chemistry having autonomy with respect to how in general 'their' students should be assessed is not self-evidently ridiculous. However, even here there are clear constraints. Three examples make this obvious. If staff wish their 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 central registry.

All of these constraints were accepted as proper. However, all interviewees had clear worries that their professional autonomy (their academic freedom) would be increasingly constrained or eroded in future. There were fears that central direction from 'the top' within the University would extend into more aspects of

assessment, that extra-institutional and non-subject based policies would increasingly 'interfere' and that much more precise compliance would be required, limiting the scope for interpretation of policy and resistance. Interviewees seemed to see a clear distinction between authoritative constraint and authoritarian control.

EVALUATIVE COMMENT FROM INTERVIEWEES

It was no part of the researcher's remit to evaluate assessment practice.

When it comes to concluding comment, it is difficult to avoid slipping into judgmental mode. It was decided to re-visit the data and attempt to report in summary form the evaluations of practice offered by the interviewees themselves.

(i) Assessment practice in chemistry was seen as highly effective because of its traditional virtues. On the other hand, it was not as good as it should be; there were desirable changes which had not happened.

(ii) People in the department were good at 'doing' assessment; however, there were some people incapable of capitalising on assessment opportunities and there were topics within assessment on which staff were not well-informed.

(iii) There was a pleasing diversity and balance in the assessment schemes for long-standing courses; nevertheless, when new courses were introduced, it was felt necessary to make quite radical revisions in assessment practice.

(iv) The quality of assessment practice came from the soundness of local subject-based thinking. This had a negative aspect; there had been a lack of openness to external influences.

(v) Structures existed to allow debate about assessment issues; there was a lack of effective communication amongst those who should have been involved in debate.

(vi) At the core of assessment lay the informed professional judgement of well-qualified academic assessors; the views of those assessed were essentially irrelevant. Too little had been done to listen to the student voice; it was troubling that no thought had been given to peer or self-assessment.

(vii) Policy from supra-departmental sources had not been generally welcomed. However, the Royal Society of Chemistry could exert an acceptable policy-influence. There were some areas in which institutional policy was thought legitimate and a few in which new policy would be appreciated.

CHAPTER 6 CHANGING ASSESSMENT PRACTICE: PHILOSOPHY

WAS PHILOSOPHY SEEN AS ONE SUBJECT AREA?

In choosing philosophy as a subject area to be explored, there was again the assumption that it made sense to treat philosophy as one subject area, i.e. as one discipline albeit containing specialisms within it. This assumption was confirmed by all interviewees as justifiable.

HOW DID PARTICIPANTS LOCATE PHILOSOPHY WITHIN THE ‘FOUR QUADRANTS’ FRAMEWORK?

As in chemistry, interviewees tended to talk about what philosophy courses should be like, rather than what was actually emphasised in honours assessment. Presented with the simple diagram, people had a great deal more to say about it than did the chemists; as one said, ‘the nature of our subject requires us to be critical, argumentative and indeed combative’.

Much comment was focused on the nature of the two dimensions, in particular whether the poles were genuine opposites. Despite reservations all were willing to place philosophy somewhere on the diagram. The average position was in the lower right quadrant, at (+1.5, -1.5). Honours philosophy assessment was agreed to put a greater emphasis on the evaluation of arguments than on the understanding of some defined bundle of key concepts and on a ‘pure’ rather than an ‘applied’ purpose. Philosophy was thus perceived as being ‘soft and pure’, but less decisively than was hypothesised.

WHAT SIGNIFICANT CHANGES HAD OCCURRED? WHY?

How students were assessed and what was seen as pleasing are described in Appendix V.

Changes at Level 1

There was a general view that assessment methods had not changed much. The one significant change was the increase in continuous assessment; the class essay now carried summative weighting.

Honours

The biggest change had been the move to more, and more diverse, summative assessment; the main features were the attaching of summative weighting to written products produced outwith examination conditions and the introduction of project presentations. Other changes mentioned less often were in the nature of examination questions and essay topics and in aggregation procedures (which now gave less scope for discussion and professional judgement).

There were two differing perspectives on the perceived change in questions and topics: some deplored the move to more stereotyped questions giving both staff and students less room for originality and creativity; others saw this as a desirable decrease in the idiosyncratic behaviour of topic-setters. ('I see a diminution in eccentricity as an increase in professionalism in both teaching and assessment.')

The reasons for change

(a) Interviewees' views

At level 1, the significant change in assessment was attributed to one single cause: faculty policy made necessary by University policy on modularisation. The change in assessment came about from a required change in course structure; there was no change in the methods deployed, but only in the timing and weighting of the essay and examination.

The picture at Honours Level was 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. The TQA was also thought to have influenced departmental thinking on what should have summative weighting; the department recognised that continuous assessment was fairer for students unable to do themselves justice under examination conditions and also that it motivated students to work harder 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 defend on

educational equity grounds and which they construed as being in line with the nature of the subject

At Level 1, extra-departmental policy 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.

(b) Reactions to the views of others.

Presented with the list of fifteen factors influencing past changes in assessment, interviewees reacted as 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

1. There was general agreement that the major influence on assessment change at Level 1 had been changes in course structure. It was inevitable that assessment

had been reconsidered in the procedures for course planning required by institutional policy. There had been no corresponding change in course structure at Honours Level.

2. The assessment of presentation skills in Honours projects was introduced to pay more attention to transferable skills. There was disagreement 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.

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.

4. People were saying not that any particular assessment change had been introduced out of a conviction that it would enhance student learning, but rather that a concern for student learning permeated all their thinking about assessment. This factor must be seen to be important otherwise staff might be charged with not caring enough about student learning.

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.

5. The role of external examiners in relation to assessment change was reactive rather than proactive. External examiners might advise a change in assessment methods, but they could not require it.

6. In responding to items about policy, it was frequently mentioned that it was always difficult to say where policy originated. Staff were unsure whether the faculty had formulated a particular policy or acted as the channel for an institutionally derived policy. Similarly, it was not clear whether the University had initiated a policy or acted as a channel for national policy.

Faculty normally played a reactive role. 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 a particular assessment method be used; it was seen as reasonable, however, that Faculty required a change in marking scales in the interests of inter-departmental comparability and intra-faculty uniformity.

The central management of the University was said to have generated policy on marking scales, letter grades, grade descriptors, methods of aggregation and anonymity in marking. They were also 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 got grades E, F or G. ('Everyone knows that you have to get a D to pass'.)

Some 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.

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

It was interesting that several members of staff thought that University policies had, on balance, produced some desirable consistency rather than an inappropriate uniformity. They feared that the balance was going to shift in future.

7. The influence of any individual enthusiast for change had been unimportant, except perhaps in relation to the introduction of project presentations. There had been no impact on assessment practice from any educational/assessment theory or theorist. Awareness of assessment trends in philosophy elsewhere came indirectly from external examiners or from experience of external examining. Two people said there had been some scouting around but ‘there are precious few ideas about innovative assessment within the philosophical community’.

8. All changes in assessment had been thoroughly discussed with students. Students had suggested, but not pressed for, changes in the methods by which they were assessed. It was 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 more insistently for assessment change in future.

9. With one minor exception, there had been no moves towards computer-based assessment.

10. In philosophy there was no subject-related professional body which could influence how students were assessed.

There were no differences in views on change and its causes related either to length of experience or to gender. The views of the philosophers were more homogeneous than those of the chemists.

WHAT FURTHER CHANGES DID STAFF (a) WISH TO SEE AND (b) THINK MIGHT BE REQUIRED OF THEM?

(a) Changes wished

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. 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, less dependence on essays and examinations, encouraging more originality within essays and reducing reliance on graduate teaching assistants. The first three are obvious enough, but the fourth needs some amplification. 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 high, but because of student perceptions that their marking was less reliable than that of experienced staff.

Within a department that had always attached great importance to teaching (under which heading interviewees put assessment), the ever-growing stress on research activity tended to force the demotion of teaching. Pressure on the time that staff could 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 expressed doubt as to whether there was the general,

departmental will to devote extra time to assessment. This was not complacency, but rather an indicator of stress within the system.

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 was again for more assessment that was informal and formative.

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 meant. Time to do this was a problem.

(b) Changes that might be required

Staff were invited to speculate about changes in assessment practice that might be required of them. 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; when speculating about future changes people talked much more of policy than they had when reflecting on past changes. The source of that policy was immaterial; in the hierarchical structure that operated, policy came through the faculty; where it started was not seen as a matter of much importance.

Two externally imposed requirements were predicted: to attach more weight to the assessment of transferable skills and to give even greater weighting to continuous assessment. Any move to greater diversity in assessment combined with more continuous summative assessment would 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 would be two potentially conflicting forces: one requiring more diversity in assessment and the other militating against it. Different policy demands, from different sources, would then require departments to try to move in different directions at the same time.

Influences such as these made some people think that it was inevitable that in future the University would 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....

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 difficult (and time-consuming) to generate items which effectively assessed higher intellectual abilities.

It was thought relatively unlikely that any supra-institutional body would prescribe specific 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 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.

The expectation that 'policy' would become more significant was confirmed in the structured part of the interviews.

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

WAS POLICY PERCEIVED AS INFLUENTIAL IN BRINGING ABOUT ASSESSMENT CHANGE?

From what has already been written, the obvious answer to the question is again 'not much in the past, expected to be more in the future'. Much of the comment made about the situation in chemistry is relevant here. There is, however, one difference.

Reviewing the two sets of data, there was in chemistry evidence of some internal wish, albeit a mild one, to initiate change in assessment practice. A 'general concern for the quality of student learning' had been a motivator for change; there were a few staff on occasion pro-active in relation to assessment practice. Pro-activity was almost completely absent within philosophy: assessment change had been reactive and what might be called creative reaction had not occurred.

There was again a general perception that policy would become more insistent, wide-ranging and directive; it was said that assessment was rising up the

policy agenda. It would become necessary for staff, somewhat reluctantly, to spend longer considering how future policy would be implemented within philosophy without doing violence to the perceived nature of the subject. There was less sense in philosophy that staff might themselves become more active in influencing the nature of the policy. (It will be recalled that chemistry had successfully resisted modularisation policy in the past; philosophy had not.) Philosophy staff were policy-recipients rather than policy actors.

Not long before the interviews took place, the philosophy department had been re-located from a site in the main building of the University to one in a rather peripheral avenue. Staff talked of this as being symbolic of the declining status of the subject in the eyes of University managers. People had felt impotent in the face of central direction and this feeling seemed to generalise to other areas. Whether this feeling of disempowerment would persist is unknown, but at the time of the research these participants were in Trowler's 'sinking' category. Policy might be disliked, but it was accepted as something that just had to be implemented; morale and job-satisfaction had decreased; there was a feeling of people 'going-under'. In the words of Burns: 'An' forward tho I canna' see; I guess an' fear.'

EMERGING ISSUES

Originality

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.

Others believed that the significant factor was that the topics set and the questions asked were more predictable and less open-ended than previously. 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 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 had been gained at the expense of student originality. This suggests that academic professionalism itself was in need of some re-definition.

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.

The integrity of examinations was seen as a powerful advantage, shared by orals, but not by essays, projects, dissertations and portfolios of work. Increased diversity in methods in assessment had threatened the integrity of the assessment process and there was a good argument to retreat to reliance on examinations – however unfortunate this might be for those they did not suit.

If student dishonesty could pose a threat to the validity of assessment, so too could assessor unreliability. Assessor bias need 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 inter-assessor inconsistency was unavoidable; this did not relieve people of the duty to explore the causes of inconsistency of judgement and to minimise it.

Considerable effort was expended to ensure that procedures were meticulously followed in the interest of fairness and to ensure the 'integrity' of the process. There was a view, however, that in the absence of full shared understanding of the meanings to be attached to assessment criteria 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.

Standards drift

There were few references to standards in relation to Level 1 assessment. There was concern about a perceived downward drift in standards at honours level. The cause of this was 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 seen as 'good' degrees. This had resulted in more generous marking of student essays and exam answers.

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 to project presentations; these took place in an end-of-term atmosphere and some markers were not confident enough about the application of the assessment criteria to mark students down.

Workload

Although all interviewees had things to say about the workload generated by assessment, no interviewee had any serious complaint. The topic is, however, labelled as an issue because this was the one area in which interviewees kept insisting that their views were not typical of philosophy staff as a whole.

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.

Assessment took a lot of time, but interviewees did not think that overall the time was excessive; it probably did not take more time than in the past; there was a wish for more formative assessment, but a general feeling that time was not available; staff would object to having to spend more time on assessment in the future. Whilst one person said she would welcome with open arms any innovation in assessment which saved time, she also said that no-one had time available to seek out or devise such innovations.

When the research was being arranged, the head of department said: 'Don't come at exam times. We are far too busy **doing** assessment to think about it'. This does not of course mean that 'doing' assessment was thoughtless practice.

IMPORTED ISSUES

Dominant purposes of assessment

There was 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.

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; they were less sure this was true of the learning function.

Constructive alignment

The concept of constructive alignment was only used by two interviewees.

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.

All the interviewees focused on what was earlier called epistemological alignment. The weakness of this as a defence of traditional final examinations again went unrecognised.

Communicating about Assessment

There was material in all the interviews that related to the theme of communication about assessment. Two matters were very clear. Firstly, as in chemistry, there was an impressive amount of collaboration over the construction of assessment instruments, particularly exam questions. Course teams did on occasion meet round a table, but more usual was the circulation of drafts for comment, increasingly 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 frequently frightful arguments about controversial dissertations’.

There was, however, a majority view that there was too little departmental discussion about concerns like reliability. This was perhaps not surprising given that staff had known each other for a long time; there had been plenty of time and opportunity to develop shared understandings about assessment criteria. 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 until they become familiar with our ways’).

Although everyone said something related to communication, two people spoke at length. 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 reaching a consensus of professional

judgement through discussion. This changed process led to decisions which were easier to defend but not 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.

The second person related the amount of discussion to the relatively low priority of assessment amongst people's other commitments.

There is really remarkably little discussion of assessment during the year. 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 talk too little about assessment and I put it on the back-burner as cheerfully as everyone else.

Criterion-referenced assessment

In the philosophy interviews, there were regular references to assessment criteria and to the desirability of making criteria explicit. Otherwise 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 the actual assessment scheme operated.

Constrained autonomy

The views of the philosophy staff were the same as those of the chemists, the only difference being there was no external subject-related body to constrain autonomy in assessment practice.

Central University prescription relating to assessment was expected to increase, to extend into new areas and to become more directive. Policy constraints would increasingly come from supra-institutional sources lacking appreciation of the needs and distinctive nature of individual subjects.

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. 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?

EVALUATIVE COMMENT FROM INTERVIEWEES

- (i) Assessment practice in philosophy was seen as effective because of its traditional character and the congruence of assessment methods with the nature of the subject.
- (ii) Staff within the department were judged good at assessment; on the other hand a great deal of assessment was carried out by graduate teaching assistants and there were questions about their capabilities. Staff had no great knowledge of possible innovations in assessment and emphasised that there was neither the time nor the will to find out about them.
- (iii) What changes there had been in assessment had come about in response to external policy influences; there had been no radical changes in courses and no

entirely new courses, thus assessment had not been re-considered as part of course planning.

(iv) The quality of assessment practice came from tradition and experience and the soundness of local subject-based thinking. There had been policy pressure, but no educational or technological input.

(v) There was plenty of communication on borderline essays and dissertations, but little enthusiasm for discussion of the big issues in assessment.

(vi) At the core of assessment lay the professional judgement of well-informed academic assessors; changes in aggregation procedures had reduced the scope for this. Students were involved more in discussion about assessment than in chemistry, but they had no greater influence as change agents.

(vii) There were some areas in which institutional policy was clearly legitimate and had been positively beneficial; future policy intervention was expected to be more malign. The philosophers would again be policy-recipients rather than policy actors.

CHAPTER 7 CHANGING ASSESSMENT PRACTICE: MEDICINE

PREAMBLE

The research focused on the assessment of student learning within the five year MB, ChB degree programme. A new curriculum came into operation in 1996; this was planned to be in line with the recommendations in *Tomorrow's Doctors*. (GMC, 1993) When the interviews took place, the first students on this new curriculum were half-way through their final year. A description of the new curriculum, its purposes and assessment, will be found in Appendix VI.

WAS MEDICINE SEEN AS ONE SUBJECT AREA?

The assumption that it made sense to treat medicine as one subject area was confirmed as justified 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?

HOW DID PARTICIPANTS LOCATE MEDICINE WITHIN THE 'FOUR QUADRANTS' FRAMEWORK?

In general the framework was viewed favourably: 'this is a good grid – it makes me think'. Again, 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'.

The vertical dimension was seen as less problematic than the horizontal dimension. The inescapable central concern at the end of the five year programme must be whether students could be judged as fit and safe to practise (under

supervision); this provided the dominant emphasis in end of programme assessment. When considering the horizontal axis, staff were hesitant about accepting the poles as genuine opposites. 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 core concepts. Final year students must be able to demonstrate their understanding of both.

Eight of the eleven interviewees were reluctant to place medicine at one location on the diagram; they insisted on two points, one for the 'old' curriculum and one for the new. 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 (applied-hard) quadrant; this was indeed the result for the old curriculum, but not for the new. There were three explanations suggested for the left-to-right shift: (a) there was less emphasis in the new curriculum on the physical science knowledge bases of medicine and more on social and ethical aspects; (b) there was now more emphasis on making use of knowledge than on possessing it and (c) there was more emphasis on informed critique of alternative theories, diagnoses and treatment strategies.

WHAT SIGNIFICANT CHANGES HAD OCCURRED? WHY?

How students were assessed in the new curriculum and what staff thought was pleasing about assessment are described in Appendix VI.

Significant changes in assessment

The form of the question was the same as in other subject areas. However, interviewees focused on the ways in which assessment in the new curriculum were significantly different. 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 continuing modifications that were made as the assessment system was implemented. The difference between 'major original' and 'minor subsequent' might not be 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 been modified.

There were four changes described as 'most significant' by the 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

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' of which they used little 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.

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'; this showed itself in the capacity to analyse clinically situated scenarios/problems and then go on to weigh up different theories, diagnoses and treatment/management strategies. The changes in assessment were interpreted as 'Bloom' category shifts: from Knowledge and Comprehension either to Application or to Analysis and Evaluation. This change was the response to the GMC's principal policy recommendation on assessment: 'Systems of assessment should reduce emphasis on the uncritical acquisition of facts'.

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.

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.

Responses were about 'integrated' assessment. However, there were distinctions and differences of emphasis. Some people talked about what might be called **integrative** assessment, where one assessment method required students to pull-together understandings from different subject areas. A good example of this was the Modified Essay Question which began with one patient-based scenario and then provided sets of questions carefully structured to draw on understanding of a range of underpinning knowledge bases. Other people described what we may call

integrable assessments; they saw the requirement within the new curriculum to relate individual assessment episodes to some organising schema. A framework 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. Staff, 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.

It was noted that the change to more integrated assessment had resulted in significant changes in who actually **did** the assessment i.e. who designed, managed and marked it.

More standardised clinical assessment

Clinical assessment within the previous curriculum was described as having been 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'.

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 brought about a big change in clinical assessment. These methods were seen as more standardised and reproducible and allowing more consistent professional judgement; in addition their introduction had required a serious exercise in the clarification of the assessment criteria. The methods attempted to increase the reliability of assessment, while maintaining authenticity and validity.

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. Staff recognised that methods of assessment gave strong messages to students about the type of learning that is required; there was a good chance that the new assessment methods 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. Students are beginning to get the message that low-level factual knowledge is just not enough.

There were two other changes seen as significant, but not by a majority. The first was increased alignment. 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. There was now a closer correspondence between the rhetoric of course purposes and the reality of assessment practice. Just one example of this was the Medical Independent Learning Examination; independent learning was said to be important and this was then confirmed by serious assessment of independent learning. The second change was a shift to more continuous summative assessment; the existence of 'necessary hurdle' assessments and coursework assessments had both reduced the stressfulness of 'big-bang', terminal examinations.

The reasons for the changes

(a) Interviewees' views

Asked what they saw as the reasons for assessment change, people in medicine had one straightforward answer immediately to hand: 'Assessment changed because the course as a whole was re-designed'. But why had these particular changes been introduced? Reasons 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 sent 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 staff said that assessment had changed to bring it into closer harmony with what students would be doing when qualified: 'How students are assessed should be strongly influenced by what they will be doing as pre-registration house officers'. There was, however, a difference between aligning assessment better with the nature of future employment and aligning it better with educational purposes. One interviewee made this very clear.

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. Merely competent doctors are rather less than professionals who have been educated for future demands.

This quotation raises profound questions about the purposes of higher education and the nature of professionalism. Within this new curriculum were students being assessed on their fitness to practise or on the extent to which they had benefited from higher education? Should professionalism be so conceptualised as to require those aspiring to it to be both proficient within existing practice and critical of it?

(b) Reactions to the views of others

Presented with the list of fifteen factors influencing past changes in assessment everyone provided crisp responses. 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

Only three factors provoked any extended 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 not the latter. Secondly, there was 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 had not simply 'fallen into line'. 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 seemed that educational theory and theorists were of no importance whatsoever. It is significant that there is a substantial body of research in medical education and assessment; it was this research literature that people quoted as influential and not the general education/assessment literature. The interviewees 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.

Many factors were seen as unimportant. The number of medical students has not changed recently, thus number of students had no influence on assessment change. Admission to medicine is highly selective, thus the student body had not become more diverse in its abilities. Although there are many different career paths within medicine, the medical curriculum did not prepare 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 increased emphasis on communication skills and their assessment, but these were seen as key skills becoming more important within medicine rather than skills transferable to other jobs.

WHAT FURTHER CHANGES (a) DID STAFF WISH TO SEE AND (b) THINK MIGHT BE REQUIRED OF THEM?

(a) Changes wished

Responses were thoughtful and rich in ideas. The general view was, 'It's pretty good, all things considered, but it could be even better'. It was striking that people did not say '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'. The evidence came only from the eleven interviewed, but this researcher was forcibly struck by the commitment to assessment improvement of hard-worked people whose practice had already gone through considerable upheaval.

Most of the recommendations for change related to three broad themes:

- giving students a better idea of where they are;
- instituting better review of assessment overall;
- further improvements in clinical assessment.

(i) Giving students a better idea where they are.

Everyone spoke of this, but there was no simple consensus about what it meant or how to do it. People frequently mentioned formative assessment: 'All the hard thinking went into summative assessment; formative assessment remains very patchy'. About half construed assessment as being formative if an important purpose was to give students useful feedback. 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 were different underlying assumptions; in the first case, the benefit of formative assessment lay in the helpful effect on future student learning of feedback from staff; in the latter case students benefited from what they learned for themselves from having a 'dry-run'.

Staff gave different reasons for wanting more 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 interesting in that they suggest three different ways of interpreting the title theme. Students should have a better idea of where they are. In relation to what? To a defined set of learning outcomes? To their peers? Or to themselves at an earlier stage? There is a pleasing parallel here with three fundamental approaches to assessment, criterion-referenced, norm-referenced and ipsative.

One person provided a somewhat different reason for more formative assessment: it would decrease 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. 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.

Related to this theme were several recommendations about desired changes in how the results of assessment were 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 asked for finer verdicts. ‘Was I only just satisfactory or securely satisfactory?’ ‘Or exceptionally satisfactory.’ ‘Was I nearly satisfactory or lamentably unsatisfactory?’

(ii) Better review of coherence of assessment

People went into less detail on this topic than on the previous one, but felt 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. 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.

Although the most frequently expressed arguments for a review group related to coherence and progression within assessment, such a group would serve other valuable 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.

(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 more tentatively than on the previous one, probably due to the not-yet-fully-tested clinical assessment arrangements.

Changes that people wished were of three types. Firstly, there should be wider sampling of clinical and practical skills within Objective Structured Clinical Examinations (OSCEs). Secondly, the range of specialisms represented in the Modified Long Cases should be reviewed and 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. Everyone wished change in clinical assessment to satisfy three criteria (a) no reduction in the authenticity of such assessment, (b) an increase in the reliability of such assessment and (c) no increase in the assessment burden falling on clinicians.

Changes that may be required

The changes that people foresaw involved extra-faculty pressure and policy, except one.

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.

Four extra-faculty policy sources were identified 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; 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 policy group on assessment. Nine of the eleven interviewed thought it possible that the University might require Medicine to use a grading scale that would be uniformly adopted across the University. Reactions to this possibility covered a whole spectrum of opinion, but tended to be negative.

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.

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.

It may be thought strange that a grading scale should arouse strong, diverse reactions. Medicine needed to operate two approaches to assessment, each with an appropriate form of grading. When assessing for clinical competence, then a student was either fit to practise or not, i.e. a simple binary decision was all that was strictly necessary. When assessing for academic understanding, then levels of achievement could be differentiated and extended point scales might be appropriate. If it insisted on a single grading scale, the University would be attempting to impose 'an unholy union of incompatibles that could only work if rationality was suspended'.

A majority 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 lay in GMC policy 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 the next major policy document from the GMC would be more prescriptive and policy imperatives would focus specifically on assessment.

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 ineffective communicators. This might be translated into a requirement for change in assessment priorities and methods. Each year a very small number of students were academically and technically sound, but unsuited to the practice of medicine because of their attitude and approach. It might be necessary to use Practice Panels which effectively decoupled graduation from registration. This would have profound repercussions on assessment.

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 assessment.

The QAA might require change. This was not because of some precept in the assessment section of their code of practice, but because of Academic Review.

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.

In the opening sentence of this section, both pressure and policy were mentioned. 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. Such 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 deriving from some body with legitimate authority. 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 then be made available to support a claim that it had been implemented in some form or another. The details of what was adopted might vary from context to context. Policy was not seen simply as a directive from on high that had to be implemented in detail; policy allowed scope for interpretation and re-construction.

The expectation that 'policy' would become more significant in bringing about future change in assessment practice was confirmed in the structured part of the interviews.

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

WAS POLICY PERCEIVED AS INFLUENTIAL IN BRINGING ABOUT ASSESSMENT CHANGE?

Policy deriving from the General Medical Council was seen as being very important in bringing about assessment changes within a re-constructed curriculum. In the past, other national policy sources and institutional policy had been 'not important for medicine'. In the future, all policy sources were predicted to become more influential.

Re-visiting the relevant data in the interviews, it was striking that some interviewees spoke of GMC policy as if it had prescribed how students were to be assessed. ('The GMC dictated methods of assessing students and we did as we were told.') Compared with what the GMC actually said, it was clear that interviewees perceived the GMC as having been more prescriptive than it actually was; the course planners for the new curriculum had a great deal of scope for creative response to policy demands. It is the case that the 1993 GMC policy language is direct, unambiguous and at times peremptory. However, its imperatives are at the level of principles: 'Assessment schemes must adequately test the educational goals highlighted in this report'; 'Assessment of the core curriculum must be rigorous'; 'Guidance and training will be required for those who do the assessing'. The GMC has little to say about methods of assessment and what little it does say is couched in the gentler discourse of advice about good practice: 'Methods of assessment will vary according to the nature of modules, but will often take the form of a short dissertation'; 'The multiple-choice format tends to emphasise the acquisition of facts at the expense of reasoning'.

Strong feelings about policy direction from the University appeared to be due to two factors. Firstly, the University seemed about to prescribe that one

grading scheme be deployed. The requirement here was not for the acceptance of some general principle but for the adoption of a uniform procedure. Secondly, the policy requirement was emerging from a non-medically based body: 'Non-medics are telling us what we as medics must do'. There is a clear resonance with the position outlined previously for chemistry

There was a sharp contrast between medical staff reaction to emerging institutional policy and that of the philosophers. In the latter case, experience of central management direction had led to a perceived decline in status and an increase in feelings of impotence in the face of current and future demands. In the medical case, a high-status faculty had perceived a threat within a draft policy document and had quickly responded with intense policy activity. This had resulted in a shift in the ways of working of the institutional policy group (to direct consultation with faculty groupings) and in its thinking (towards groups of faculties selecting the grading scheme they felt appropriate to their own interests). If the philosophers were 'sinking', the medics were fighting vigorously to have policy re-defined to suit them.

EMERGING ISSUES

The appropriate grading scale

This issue was not merely emergent, or even salient; it was irruptive. A fuller treatment will be found in Appendix VI, pp 30-32. At the heart of the issue is whether, for medical students, a piece of assessed work (or observable performance) should be graded as (a) yes-no, (b) merit, pass, fail, (c) highly satisfactory, definitely satisfactory, just satisfactory, unsatisfactory, (d) A to E, (e) 20 to 1 or (f) a percentage.

In the old curriculum, a range of grading procedures had continued unquestioned. In the new curriculum, course-planning procedures had prompted some simplification. The original planners had strongly favoured simple pass/fail decisions for all course assessments. As a result of the first years of the course, sub-groups within the faculty began arguing for different grading systems. When it looked as if central University policy might require the adoption of a single scale, the medical faculty (although divided within itself) united against an external imposition which would be 'the worst of all possible worlds'. Vigorous policy activity resulted and this brought about a shift in the working and thinking of the University policy-generating group. This did not go as far as allowing the medical faculty to do whatever they liked. The special character of courses of professional preparation was agreed to justify a different approach to grading from that within 'academic' subjects. This, however, implied that there should be uniformity in the grading scales employed in medicine, teaching, social work, law; the last three of these are not groupings the medical faculty ever saw any need to co-operate with in the past. At the time of writing the issue is still being addressed, most vigorously within a new Assessment Review Group set up within the medical faculty.

How is this to be summarised in terms of policy and practice? Within medicine, assessment practice changed radically as part of a general curriculum change; the main prompt for this was extra-institutional policy generated by a subject-specific professional body. Soon after this, institutional policy was generated which might have had a direct impact on assessment practice in medicine. When this was realised there was an upsurge of policy activity from medical staff which (a) tended to preserve medical faculty autonomy and (b) had a significant influence on institutional policy-making.

Workload shifts and the management of assessment

The new assessment arrangements required a great deal of work from people, but significantly these were not the same people as 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.

Different people, with different titles, in different places, were seen as doing the work and inevitably this meant that different 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.

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. Authority is a difficulty; I have to rely on goodwill.

There was a lot of goodwill around; the general view was that ‘people are usually helpful if you say please’. However, there were some people who were now unwilling to get involved with assessment; 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.

The expertise needed by assessors

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. 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.

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.

There were two disturbing features in the interview data. Firstly, although several members noted that they wished to learn more about assessment, only one person spoke of any action 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.

It would appear that para 56 of the 1993 GMC Report had been largely forgotten:

‘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.’ Secondly, there was too much knowledge of assessment for any one person to possess, so there should be effective pooling of the wisdom and knowledge that **does** exist. The single most dispiriting remark in the interviews was, ‘In medicine there are still A-list people and B-list people; the As don't listen to the Bs’.

Authenticity and reliability

A quotation illustrates the core 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?

Most saw this as an important issue, but people talked about it using different language and concepts. Three people, for example, talked of problems in reconciling ‘reality’ with consistency. (Using ‘real’ patients was a good thing, but in their infinite variety they made it difficult for assessment to be consistent and fair to students.) Another three discussed the issue in the traditional language of validity and reliability. (‘In the new curriculum we stressed validity and then had a problem with reliability’.) An important point was made – that medical teachers press for

validity as an educational obligation, whereas administrators seek reliability (because they face the brunt of student appeals based on perceived unfairness in unreliable assessment.)

The contribution of one person was memorable. She preferred to speak of authenticity and reliability rather than validity and reliability. Conventionally people had sought reliability and then added on validity an afterthought, if at all. She preferred to de-couple the two ideas and abandon the standard thinking on their relationship (i.e. that reliability was a necessary but not sufficient condition for validity). It was more sound, she said, to go all out for authenticity and then try to make assessment as reliable as it could be given the practical realities and resource constraints. When asked if she had been influenced in her thinking by anyone in particular, she immediately provided the precise reference (van der Vleuten, 2000). In other subject areas, there was no reference to **any** literature on assessment.

It seems only appropriate to conclude with a quotation from that interviewee.

There is reliable assessment which is quite cheap – paper and pencil MCQs. There is authentic assessment which is quite cheap – observing students with patients. Assessment which is both reliable and authentic does not come cheap. If we want better assessment, then the money has to be found to pay for it.

IMPORTED ISSUES

Dominant purposes of assessment

There was little reference to the accountability purpose. One person noted that more use could be made of assessment data in course evaluation; another noted that assessment processes carried out by facilitators with their groups provided them with feedback on their own performance.

Everyone clearly acknowledged the certification and learning enhancement purposes; both were seen as inescapable. Although there was concern about the 'rigour' of clinical assessment in the final years, there was no fear that the qualification awarded would have lower currency. There was general agreement that the new assessment scheme did more to enhance student learning than had happened in the past. However, the number of comments about giving more attention to formative assessment suggests that people were aware of the scope for fulfilling the growth function better.

Constructive alignment

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 was thus not 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 design principle. What surprised the interviewer a little was that the actual term was used by interviewees. What might be seen as a recent coinage, or an unwelcome bit of jargon, had become for some medical staff an unremarkable part of everyday language usage.

Communicating about assessment

There were three areas in which staff were said to communicate very effectively. Firstly, there was the MILE; secondly, there was the Longitudinal Care Project (in which there had to be much communication amongst University staff and

general practitioners); thirdly, there were the termly discussions amongst facilitators and block-leaders about student progress.

Sometimes communication was less than ideal. 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 and 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'.

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'. The need for purposeful communication was related to the view of most interviewees that the Assessment Review Group should be re-activated. There was a clearly stated wish that discussion about assessment be promoted up the list of faculty priorities and not be perceived as a trivial, social matter.

Criterion-referenced assessment

What has already been written about grading scales could be interpreted as a conflict between criterion-referenced and norm-referenced thinking. The debate could be seen as people trying to think their way through to a workable compromise between conflicting ideologies. There appeared to be two related ways in medicine along which people varied. Firstly, people thinking administratively preferred

students to be spread out (N-R A); people thinking educationally wanted most students to be successful and a few clearly unsuccessful (C-R A). Secondly, some could see clear differences in the levels of academic understanding that students achieved and thus a justification for differentiating students across a spectrum of grades (N-R A); others could not and thought in terms of the binary distinction between competent and not-yet-competent. (C-R A). It may be that the Medical Faculty may agree to using a University scale to spread students out in 'academic' matters and modified-binary decision-making to report on clinical assessment.

Constrained autonomy

Constraints on individuals and the faculty were again seen as proper and reasonable. 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 extend 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 seen as less unwelcome when it came from people with 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 suspicion because 'they tend not to appreciate our distinctive character'.

EVALUATIVE COMMENT FROM INTERVIEWEES

- Assessment is much better in the new curriculum, but further improvement is desirable.
- There is need for an assessment review group to ensure assessment issues are regularly discussed.
- New forms of assessment need more communication and changed management.
- Not all assessors possess the assessment expertise required.

CHAPTER 8 CHANGING ASSESSMENT PRACTICE: DESIGN

PREAMBLE

Interviews in all four subject areas were productive, interesting and enjoyable. If the researcher were pressed to identify the subject area that was most productive, in terms of data relevant to his research purposes, it would be medicine. If pressed to name the subject in which the work was most interesting and enjoyable, it would undoubtedly be design. There were four reasons for this. Firstly, the area was new to him; there was much to be learned. Secondly, as someone who cares passionately about the quality of student learning experiences, he was delighted to meet staff who shared this passion and who operated an assessment scheme which seemed compatible with it. Thirdly, this was the only subject area where he met the students themselves: the rooms in which the interviews took place were usually next to design workshops; he was taken on a tour of several design areas; interviews were occasionally 'interrupted' by students urgently needing to talk to a staff member; at the end of some interviews he talked with students. Fourthly, design staff provided good quality coffee, sometimes accompanied by Fry's creams.

The interviews were longer than in other areas, but they produced less material directly relevant to the concerns of this research. The reason is simple: there had been little change in assessment procedures during the period of which the interviewees had experience and there was little change seen as desirable or as likely to be required. The research had no evaluative purpose, but it is an indicator of how pervasive has become the concern with change that the researcher realised in the early design interviews that he was at risk of making a judgement that 'little has changed, thus it can't be very good'. To minimise the risk, he wrote an informal

paper entitled 'The Desirability of Change'. An abbreviated part of this appears in Appendix VII, pages 23-25. This concludes as follows.

'The position of the design staff can be summarised as follows:

- (i) our assessment procedures were very good in the past;
- (ii) there has been little change in the context of our operations that impinges significantly on assessment;
- (iii) there has been little significant change in our assessment practice;
- (iv) our assessment procedures are still good;
- (v) there is no pressing need for much future change in our assessment practice;
- (vi) given more time for assessment we could do what we currently do but rather better, and also make minor evolutionary changes we recognise as desirable.'

The following quotation illustrates this 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.

How students were assessed, and what interviewees saw as good, are described in Appendix VII. Within all assessment of studio practice there were four important features. Firstly, portfolio assessment was central and crucial. Secondly, all assessment at all stages made use of the same broad assessment criteria, each sub-divided into five or six varying sub-criteria. Thirdly, all design courses had both interim and final assessments, informed by an 'Assessment Grid' embodying the criteria and levels of capability appropriate to stage of programme. Fourthly, all years followed submission of the portfolio for degree decision-making with a public showing of student work.

There was a wide measure of agreement about what was good in the assessment procedures. Firstly, all assessment was based on dialogue with students. Secondly, the criteria/levels grid reflected the nature of the design process and helped to make the whole assessment process authentic, coherent and progressive. Thirdly, any potential unfairness to students arising from unavoidable subjectivity in design assessment was mitigated by a large amount of collaboration amongst assessors. Fourthly, the public event of 'The Degree Show' prompted a regular appraisal, with a wider audience, of the meanings attached to assessment criteria.

WAS DESIGN SEEN AS ONE SUBJECT AREA?

The assumption that design could be treated as one subject area was confirmed in the interviews, although within a broad common purpose and much shared assumption and language, there were clear differences from one 'discipline' to another and indeed within them. The fact that there was a School of Design at all was given as evidence of belief in a general concept of 'design'.

HOW DID INTERVIEWEES LOCATE DESIGN WITHIN THE 'FOUR QUADRANTS' FRAMEWORK?

Everyone had a lot to say. 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 describe what the design curriculum **should** be like rather than what was actually emphasised in final year assessments. It seems likely that in design there is a genuinely close correspondence between the rhetoric of course purposes and the reality of assessment practice, because of the centrality of portfolio assessment and the high visibility of portfolio contents.

The most frequent comments 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 one particular career; (b) the different disciplines within design would be located differently and historical/critical studies would be different from studio practice; (c) understanding of concepts and evaluation of alternatives are not polar opposites, they proceed in parallel. Interestingly two people offered alternative frameworks which they thought better captured the situation existing within design.

When it came to locating design on the diagram, people fell into two main groups. A majority placed design firmly in the upper right quadrant ('applied and soft'); the minority said that design had to embody both sides of both dimensions, thus it could only sensibly be depicted within a shape centred on the origin. A crude average for all the design interviewees located design in the upper right quadrant, (at +1.5, + 2.0). This represents a perception that in the final assessment of design students, more emphasis is placed on their fitness to practise than on their academic soundness and more on their ability to evaluate alternatives than on their understanding of a bundle of defined core concepts. Design was chosen for this research in the hope that design teachers would locate their subject in the upper right quadrant; this they did.

WHAT SIGNIFICANT CHANGES HAD OCCURRED? WHY?

Changes in assessment practice

It was the general view amongst interviewees that there had been little significant change in assessment practice during their time as assessors of design students at GSA.

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 change, but all interviewees did identify some changes that had occurred. The most frequently mentioned changes were in the assessment criteria.

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. Firstly, increasing attention was paid to the development and assessment of oral and written communication skills. Secondly, there was a growth in the awareness of staff of the importance to learning of student self-assessment. Thirdly, there had been the welcome introduction of a choice in the weighting of the written work on Historical and Critical Studies in the final year. Fourthly, there was now greater insistence that assessment decisions be well-evidenced; assessment practice not only had to be fair, but it had to be possible to support its fairness if 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).

One person drew a very 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 increased 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. 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 not changed dramatically. There had been a number of changes in the main assessment categories and their constituent criteria and in the ways these were represented. Although the two areas had 'departed somewhat' from the others, they had not separated off and severed lines of communication.

Reasons for the changes

Interviewees' views

In design areas, other than product design and product design engineering, there were three explanations offered for assessment change. Firstly, there had been 'gentle pressure' to encourage the sharing of good practice within the design school; secondly, listening to students and being responsive to their opinions had resulted in

some changes in assessment criteria; thirdly, 'policy documents' from the Quality Assurance Agency had brought about changes in the language of assessment, but not in any central matter of assessment procedure. In the other two design areas, assessment had changed because of the need (a) to re-think it as part of the more general process of course design and (b) to co-operate with new partners within both the University of Glasgow and European institutions.

Reactions to the views of others

Presented with the list of factors said to have influenced changes in assessment, interviewees reacted as 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

By this time in the interview, everyone was giving thoughtful responses; their elaborations prompt the following comments.

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 was of three or four people, more interested in assessment than the others, who had no great difficulty in encouraging others to 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; 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 pressure for change. There are different ways in which this can be interpreted. 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 provided a picture of people

with considerable confidence in their assessment practice; there was no evidence that such confidence was ill-founded.

Staff were not isolated 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 assessment. 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 CNAAs 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 influenced assessment practice.

WHAT FURTHER CHANGES DID STAFF (a) WISH TO SEE AND (b) THINK MIGHT BE REQUIRED OF THEM?

Changes wished

People had to think hard to come up with suggestions about any change in assessment they wanted to see: 'We're really quite happy with assessment the way it is'. It was notable how many people made some reference to **time** in their replies. Given more time, however, people would not do anything very different; they were seeking a change in conditions rather than any radical change in assessment practice.

There were only five suggestions for any definite change in assessment practice. Firstly, there should be greater comparability of levels and standards across the five specialisms of the BA (Hons) Degree. Secondly, action was required to reverse a perceived drift in standards. Thirdly, in Historical and Critical Studies

there should be more scope for student negotiation of non-standard assignments for assessment. 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 would take considerable time, expertise and goodwill from all 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 will 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'.

Changes that may be required.

The general answer 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 were given more than a single mention.

The first was that an undesirable degree of conformity of practice across the separate design departments might be imposed by school or institutional policy.

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 experience of modular curricula and a general distrust of any assessment involving numbers ('the wholly spurious validity which some people attach to meaningless numbers').

All the other changes identified were 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 initial selection of students 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 likely sources of pressure for future change were given as the School of Design, central management in the Glasgow School of Art, the Quality Assurance

Agency and what might vaguely be called 'the government'. This was confirmed in responses to 'the fifteen factors'.

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

EMERGING ISSUES

Valid assessment of creativity

This is expressed as one single issue, but it could well be deconstructed into a range of 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. All that can attempted here is a brief organisation of the views of interviewees.

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 all 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

reproduced; 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 borrow something. There is little that is wholly new.

Students of design have to generate something of their own. The situation is, however, different from that in Fine Art. In design, the thing created has to meet a complex set of practical demands stemming directly from the initial design brief; the judgement then required is not merely of the design's 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 purpose. Such synthesis, although remaining 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.

How then was creativity recognised? It intrigued this writer that creative synthesis was labelled by one person as a competence. There was an implication here that by calling the creative act a competence, its recognition seemed less problematic than it actually was. Undoubtedly there are within design many skills 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 interviewee observed that Category V behaviours within the Bloom taxonomy (i.e. 'synthesis') were 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 worked 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 institutional status.

The appropriateness of honours degrees

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: the 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.

When there was such broad agreement about the undesirability of classified honours, why had they 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.

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 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 still very difficult to innovate. 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.

IMPORTED ISSUES

Dominant purposes of assessment

There was not much 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 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 supported this claim rather more convincingly than in the other subject areas.

Constructive alignment

Only relatively recently have policy-required approaches to course design emphasised that assessment procedures must be an essential 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 was very obviously present in the contribution noted on page 169.

What was clearly revealed in all interviews was the existence of discipline-appropriate alignment. Assessment was planned and carried out in ways closely and explicitly related to the perceived nature of design activity. The underpinning rationale emphasised authenticity (assessment of students closely resembled the ways in which qualified design practitioners were themselves judged) rather than validity (assessment of students allowed sound inferences about the achievement of explicit course purposes).

Communicating about assessment

The assessment procedures deployed within design were characterised by co-operation and collaboration. Standard assessment practice in design meant that assessment had become something 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 geography of GSA buildings means that we talk about student assessment not just around tables, but as we walk along the pavements.

Criterion-referenced assessment

Of the four subject areas studied, the two concerned with professional preparation (medicine and design) had moved significantly further towards criterion-referenced thinking 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 attention to 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 an intriguing, perhaps disturbing, reversal of convention.)

Within the interview data there was also a distinct emphasis on ipsative assessment (where the main concern was assessing one student against that same student's earlier performance). The assessment profile document had been designed with this specifically in mind and provided good illustrations of student development. There was also a type of assessment thinking that appeared in the contributions of two people that was unusual and deserves a new coinage, 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.

Constrained autonomy

The existing constraints on individuals and on the design school were seen as entirely proper and reasonable.

Within the school, there seemed to be acceptance of specialism-based differences in aspects of assessment practice within one broad approach; there was some anxiety, but not very much, about 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 the 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.

EVALUATIVE COMMENT FROM INTERVIEWEES

- (a) Assessment in design emphasised authenticity, dialogue and growth in a distinctive way. In design, assessment was seen as normally occurring **during** teaching/learning and not after it.
- (b) Only where new degrees had been introduced had there been any significant change in assessment and such change was not about studio practice. The absence of radical change should not imply there was need for it.

- (c) Minor changes in assessment were due mainly to internal factors; external policy had played a very minor role.
- (d) There was unease that, in future, policy requirements would mean a move to more uniformity than was desirable.
- (e) Practitioners had an informed understanding of how assessment affected student learning. They had enough time to do assessment, but not to reflect on their practice or to plan and implement innovation.

CHAPTER 9 ANSWERING QUESTIONS AND DISCUSSING ISSUES

INTRODUCTION

This long chapter is divided into six sections. These are listed below with an indication of where answers to the research questions (RQs) can be found.

Section 1	Locating subjects within the four-quadrant framework.	RQ 8
Section 2	Changing assessment practice	RQs [3], 4, 5
Section 3	Relating assessment policy and practice	RQs 1, 2, 6
Section 4	Issues emerging and imported	
Section 5	The impact of policy and practice on professionalism	RQ 7
Section 6	A personal reflection on salient issues	

SECTION 1 LOCATING SUBJECTS IN THE FOUR QUADRANTS

The four subject areas were originally chosen in the hope that one subject would be located by participants in each of the four quadrants of a framework formed by two axes ('hard-soft' and 'pure-applied' – see Fig 1, Page 8). How interviewees reacted to this framework and how they placed their subjects have been described in Chapters 5, 6, 7 and 8. The locations are summarised in the figure below.

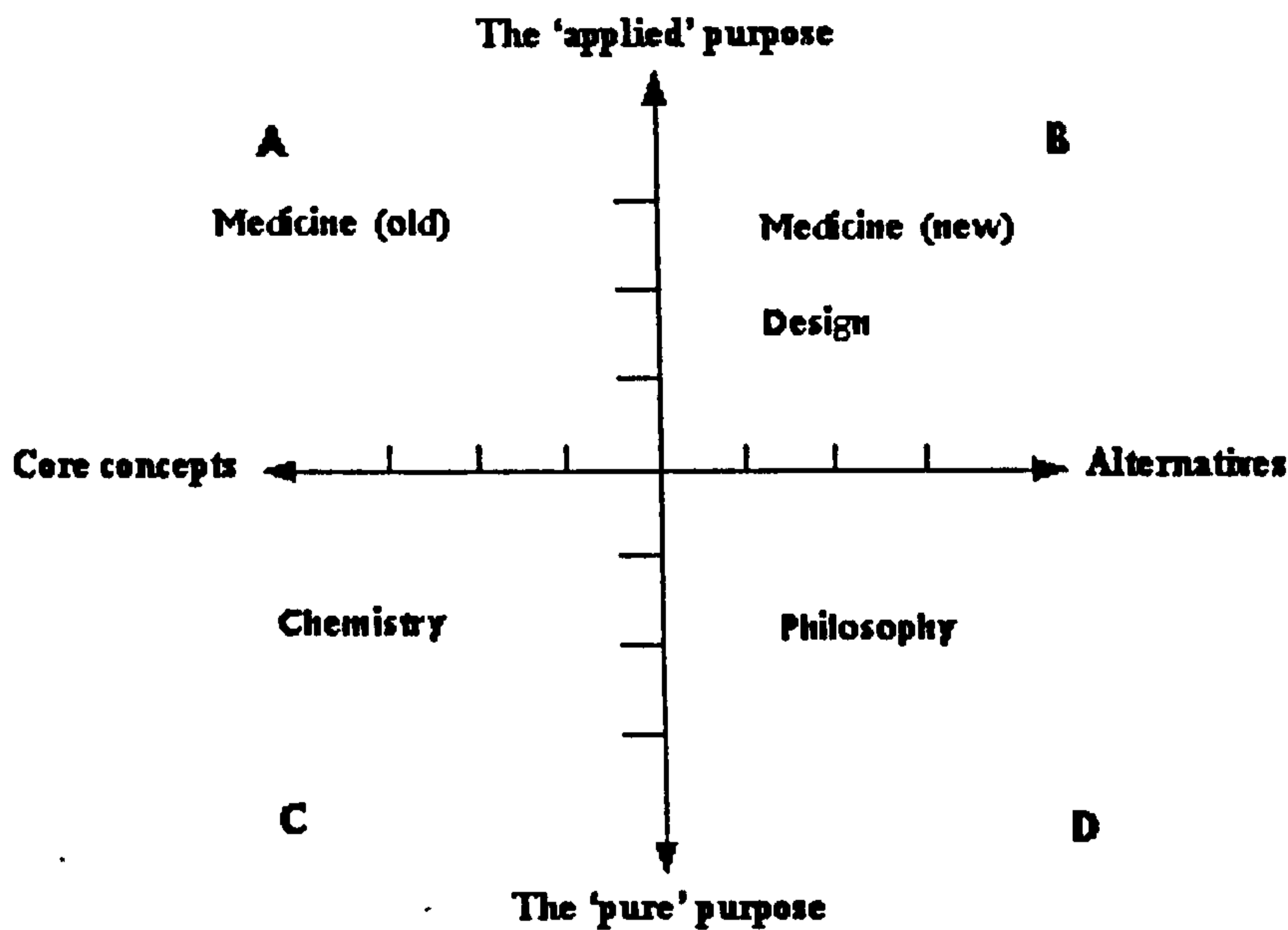


Fig. 3 Subject areas as located by interviewees on 'four quadrants' framework

There are six points to be made.

1. There was a wish to conduct the research in subject areas distinctively different in how staff thought about and practised assessment. The framework guided the choice of subject areas. The research data indicated that the chosen subjects were indeed distinct and in ways the framework predicted; staff located their subjects in the quadrants as hoped.
2. The hard-soft and pure-applied dimensions were interpreted for the research in terms of what staff sought when assessing students towards the end of degree courses. Thus 'hard' was presented as 'seeking evidence of student understanding of a set of core concepts' and 'soft' as 'seeking evidence of student ability to evaluate alternative arguments or procedures'. 'Pure' was 'seeking evidence of academic soundness' and 'applied' was 'seeking evidence of fitness to practise'.

The majority had minor reservations about these perspectives on assessment. Nevertheless, they saw the axes as identifying valid distinctions between emphases

placed on different assessment functions. A pleasing number of interviewees said they found the framework useful in thinking about their assessment practice.

3. It was noticeable that staff tended to talk about what their courses and the assessment of them **should** be like, rather than what they actually emphasised in end of degree assessments. Confirmation of the correspondence between rhetoric and reality was outwith the remit of this research.

4. The responses of medical staff were particularly interesting (see Chapter VII). Eight of the eleven staff were unwilling to locate medicine in only one place. They insisted that the old curriculum had been firmly in the 'hard-applied' quadrant, but that the new curriculum had shifted assessment significantly into the 'soft-applied' quadrant. The explanations for this left-to-right shift were (a) less emphasis on assessing physical science knowledge (hard) and more on social and ethical aspects (soft), (b) more emphasis on making use of knowledge rather than possessing it and (c) more emphasis on the critique of alternative theories, differential diagnoses and alternative treatment plans.

It should be noted that the second of these explanations hints that the two axes may not be genuinely orthogonal and independent. It was not easy to distinguish between using concepts in the evaluation of alternative arguments/procedures (the horizontal axis) and applying knowledge in professional situations (the vertical axis).

5. It is necessary to ask what these locations are and indeed whether the suggested differences between them were significant. The co-ordinates of each location are crude arithmetical averages reached from the decisions of a set of individuals. It would be possible to apply tests of statistical significance to show that between group variation was significantly higher than within-group variation.

This was not done because of a fear of giving them a spurious appearance of validity through statistical procedures. There is an opportunity here for further research, on greater numbers, of a more positivist and quantitative kind. The claim being made here is a modest one with some plausibility and a limited amount of descriptive validity: the data available indicated that the groups of staff perceived the purposes of their assessment procedures differently in the ways suggested.

The responses from individuals clustered around the points shown in the diagram. Those who were unwilling to place their subject clearly in one quadrant were mostly those who insisted that 'our assessment is about all four things, therefore I place it at the intersection of the axes'. If these people had been excluded from the averaging, the summary positions would have been more decisively different than those shown.

6. The results of this part of the research confirm, if such is needed, that it is far from fanciful to think of academic staff as being members of 'tribes'. (Becher, 1989; Becher and Trowler, 2001) Members of a tribe are not of course identical in anything, and certainly not in their thinking about assessment.

The kind of general claim made above was required by the purposes of the research and arose from the mode of analysis adopted. Large amounts of data were provided by individuals; the analysis was carried out to reveal similarities and to identify themes. Unsurprisingly the outcome was a generalisation modified by some mention of those who did not obviously conform. Further research suggests itself: 'The group shared similarities, but in what important ways did people differ and why?' This kind of question has recently been addressed in Australia by Samuelowicz and Bain (2002) They concluded that academic staff came to different decisions about how to assess their students on the basis of different values

and priorities attached to different educational ends. Their research assumed that individuals were free to assess their students in ways personally congenial to them, unconstrained by departmental and institutional policies, departmental and subject traditions and culture. One can imagine forms of programme and departmental organisation that would allow scope for such independence of action, but they did not exist in any of the sites explored in this present study. This did not mean that there was no possibility of an individual influencing assessment practice or working towards some innovation; they could, and they did, although much less than in the Australian study.

SECTION 2 CHANGING ASSESSMENT PRACTICE

Past changes

Interviewees were asked what significant changes in assessment practice they had witnessed. Details of their answers are in Chapters 5 to 8 and related appendices. Analysis first suggested that changes of three types were being identified; this categorisation was later refined by sub-dividing the first category. Each type of change was associated with a different pattern of causal factors perceived as influential. The four types were labelled evolutionary trends, policy-related shifts, within-course innovations and new-course introductions.

Evolutionary trends were usually introduced by words like, 'Over the years there has been a gradual shift of emphasis so that we have more of assessment [X]'. Participants did not identify particular events occurring at specific times. When giving reasons for such changes, people talked of educational/assessment fashion, moving with the spirit of the times and of conforming with the expectations of 'the folk up the hill' (i.e. faculty, the University, central management).

Secondly, there were policy-related shifts. These were not just general trends over time; people referred to particular events happening at definite times. For example, 'We had to change how we assessed students at that time because of the move to modularisation'. When giving reasons for such shifts, people identified official policies which had direct impact on assessment practice.

Thirdly, there were within-course innovations. These were new methods of assessment introduced into existing courses; the course context was not new or radically revised. This type of change was distinctively different from the previous two in that the impetus for change came from **within** the subject area and was not obviously a response to any external pressure. The mechanism seemed to be that one or two people thought that assessment could be improved, persuaded colleagues of the merits of their case and then introduced the change after obtaining approval from faculty, external examiners and any relevant subject body.

Fourthly, there were new-course introductions. In two areas, new degrees had been introduced (for example, the MSci in chemistry and the MEDes in design). In a third area, medicine, the degrees awarded remained the same but a completely new curriculum had been introduced. As part of the policy-required processes of course design it had been necessary to re-think assessment. There was a fundamental change here. In the past, assessment arrangements had normally been decided after course-planning that focused almost entirely on content coverage; assessment was not integral to that planning, but a bolt-on consideration at the end. More recent requirements on how course planning must occur had emphasised the necessity of specifying course objectives (or learning outcomes) and the ways of assessing achievement of these as necessary components of the design process.

	Chemistry	Philosophy
Evolutionary trends	<p>More continuous assessment More formative assessment and remediation More rigorous assessment of practical work More weight to project reports</p>	<p>More continuous assessment More weight to project reports Less creative essay topics and exam questions</p>
Policy-related shifts	<p>Increase in summative continuous assessment More formulaic aggregation</p>	<p>Increase in summative continuous assessment More formulaic aggregation Project presentations – communication skills</p>
In-course innovations	<p>Short diagnostic tests</p>	
New-course introductions	<p>Diverse ‘non-traditional’ methods Collaboration with new partners</p>	
Perception of change	<p>‘We haven’t tinkered with assessment, because it seems to work well.’</p>	<p>‘Assessment remains traditional, subject-congruent and effective enough.’</p>

Table 1 (a) Past assessment changes in chemistry and philosophy

	Medicine	Design
Evolutionary trends		Clarifying of assessment criteria Growing emphasis on self-assessment
Policy-related shifts		Assessment of communication skills Increased evidencing of assessment Improved documentation
In-course innovations		
New-course introductions	Less assessment of factual knowledge More integrated assessment More standardised clinical assessment Use of 'non-traditional' methods Increased alignment with purposes Less emphasis on terminal exams Collaboration with different colleagues	Integrating assessment of different components Collaborative assessment with new partners
Perception of change	'Assessment of the new curriculum is completely different.'	'Assessment has not changed fundamentally - there was no need.'

Table 1 (b) Past assessment changes in medicine and design

Assessment changes of these four types are summarised in Tables 1 (a) and (b). The fifth row in these figures gives the summary verdicts of the interviewees on the extent of change.

There are five matters that seem worthy of comment.

(a) This picture of assessment change can be compared with that given in the assessment literature. Texts on assessment give different accounts of how assessment practices have changed over the last two decades or so. {See, for example, Biggs (1999a – Chapter 8), Brown and Glasner (1999 – Chapter 2), Freeman and Lewis (1998, p 310), Toohey (1999, Chapter 9)} A conflation of the different lists is given by Holroyd (2000). The components of his list will now be given, with alongside each an indication of the subject areas in which that change featured in the current research.

Increasing emphasis on the learning enhancement purpose	Chemistry, philosophy, medicine design
Increased attention to formative assessment	Chemistry, medicine (?), design
More emphasis on criterion-referenced assessment	Chemistry (?), medicine, design
More provision of descriptive comment and feedback	Medicine, design
Diversification in assessment methods, reducing emphasis on end-of-course exams	Chemistry, philosophy, medicine design
Less reliance on assessment by teaching staff alone	Medicine, design
Assessment integral to teaching, rather than occurring after it	Medicine, design

Assessment texts give the impression of much more assessment change in higher education than had occurred in three subject areas; medicine was very different.

(b) An inventory of assessment changes in Scottish higher education (ASSHE, 1996), leaves readers with the impression of extensive and exciting innovations in all subjects and all institutions. This could be the case but it is not a conclusion that can be reached from the ASSHE evidence. In the nature of things, a survey of innovations does not reveal situations where there has been little change.

Caution is needed, because one cannot make any empirical generalisation from the present study of four subjects in one institution to that whole institution, or to the four subjects in all institutions, let alone to all subjects in all institutions. Nevertheless, in this research there was only one clear example (in chemistry) of an assessment innovation introduced into an existing course as the result of an initiative by people within that subject area. Of course there was more change than this, but it was change that was responsive and reactive (to the pressures of fashion and policy) rather than creative (stemming from individual academics being proactive). There were examples of people responding creatively to external pressure, very obviously in the case of medicine but also in chemistry and design. Such creativity in response, however, only came about when people were required to re-think their assessment practice as part of the planning of completely new courses.

(c) The most striking examples of change in assessment practice were in category four, the new-course introductions. In all the new chemistry and design degrees and in the new medical curriculum, there were striking assessment changes: diversification in methods, markedly increased alignment of assessment with

learning outcomes, more stress on integrating assessments and collaboration with new partners. There was also more consideration of the student role; students were seen as not merely those who had assessment done to them. The influence of feedback on learning, the significance of formative assessment and the place of self-assessment all became more prominent.

Where a course had been in existence for a long time, course providers seemed to believe that assessment within it had evolved to a satisfactory state and there were unconvincing grounds for innovation. This may be a justifiable stance; but some people in some subjects seemed to be perilously close to what a medical interviewee described as ‘the pathologies of practice that develop within unchanging systems’. When there was a requirement that a whole new course be devised, then academic staff proved themselves willing to re-examine assessment in creative ways and seemed to be energised by the experience.

(d) Labelling one category as ‘policy-related shifts’ risks an inference that externally-derived policy had been less influential overall than was actually the case. The factors influencing evolutionary trends were not explicitly identified as ‘policy’, but rather as pressures of fashion and of expectations from those with management roles. It would be a mistake to construe policy only as a formal directive requiring particular actions; policy in the past may well have acted to sensitise staff to particular issues and to encourage the prioritising of some concerns over others. There was also a sense in which ‘new-course introductions’ were also policy-related. Directions as to how such courses had to be planned came from supra-institutional sources, channelled through the University. If it had not been official policy that course design be conducted with assessment as an integral

component, there would not have been the new-course introductions within assessment that are such a clear feature in Table 1. We should distinguish between ‘policy requires you to implement these changes in assessment’ and ‘policy requires a process of design which requires you to re-think assessment’.

(e) In three subject areas, the summary view of interviewees was that assessment had not changed very much (‘and that’s fine’) and in the fourth assessment had changed dramatically (‘and that’s fine too’). Re-visiting the data from the not-much-change areas, there was a clear difference in the confidence with which the ‘situation satisfactory’ views were delivered. Chemistry and design were secure in their judgements: ‘the assessment system works well’. The nearest the design staff got to any expression of unease was they were ‘not as far ahead of the game as previously’. The chemists felt that their confidence was justified by the approval given by the RSC to their assessment procedures. There was no equivalent body in design, but they felt their assessments were validated by the approval of the wider design community and the public in general through degree shows.

Philosophy staff were defensive: ‘We think our traditional assessments remain effective because they seem in line with the nature of our subject’. However, there were ‘precious few creative ideas about assessment within the philosophy community’. There was no subject-related professional body equivalent to the RSC or GMC and there was no obvious way in which the larger community supported confidence in the assessment of philosophy graduates. Interestingly, the philosophers looked to the Learning and Teaching Support Network to provide useful guidance and support in philosophy assessment; one can see this organisation as made palatable by having subject-specific branches.

It is important not to leap from an estimate of the extent of past change to a judgement about the quality of present provision. Novelty is no better an indication of quality than is tradition. It was not part of this research to evaluate current assessment practice, although some description of how practitioners evaluated their own practice was possible. This research does not allow judgements of the type 'There has been little change in the assessment of philosophy students, therefore current practice is not good' or 'There have been great changes in how medical students are assessed, therefore current practice is good'.

After being invited to give their reasons for past assessment changes, interviewees were asked to react to a collection of fifteen factors culled from other sources. To assist with the interpretation of their views, Table 2 presents responses differently from the separate Chapters 5 to 8. The fifteen factors are now grouped: the first four are explicitly policy-related, the next four are labelled 'educational', the next three are directly student-related and the remainder are 'other'. (This classification is open to criticism; for example, 'general concern for the quality of student learning' is both an educational reason for change and one which is student-related.) For each factor there is an indication of the level of importance attached to it in each subject area. (In medicine, the new curriculum was introduced faculty-wide and faculty policy was thus not seen as an external source of policy influence.) Table 2 could of course be presented with numerical rankings; this might prompt voluminous detailed comment on fine differences in the numerical rankings. This would be foolish; only large differences are likely to have meaning.

Influential factors	Chemistry	Philosophy	Medicine	Design
National policy	Some	Some	No	No
Institutional policy	No	Some	No	No
Faculty policy	No	Some	N/A	No
Policy from subject-professional body	Very	No	Very	No
General concern for learning	Very	Some	Very	Very
Impact of educational theory	No	No	Some	Some
Persuaded by enthusiast	No	No	Some	Very
Assessment trends elsewhere	Some	No	Some	Some
Increased number of students	No	Some	No	Some
Increased diversity of students	Some	No	No	Some
Pressure of student opinion	No	No	No	No
Change in course structure	Some	Very	Very	Some
Concern for transferable skills	Some	Some	No	No
External examiners	Some	Some	No	No
Assessment technology	No	No	No	No

Table 2 Perceptions of importance of factors influencing past assessment change

(a) Responses were quite different in the four subject areas; again there were distinct disciplinary differences in assessment thinking. On only two factors was there some agreement that they had high importance. The first was a general concern for the quality of student learning. (People were perhaps unlikely to pronounce this **unimportant**.) The second was the influence of new course structures; the importance attached to this agreed closely with interviewees' unconstrained responses.

There was considerable agreement about what was of **no** importance in relation to past assessment change: institutional policy, pressure from students, awareness of appropriate assessment technology. Staff have probably underestimated the **indirect** influence of policy. They were not unconcerned with student opinion in all matters, rather they perceived assessment as a 'reserved matter' for academic staff. The verdict on assessment technology, it has to be confessed, caused despondency in the researcher. If assessment technology had been actively considered and then rejected this would have been one thing; however, amplifying comments showed that most staff were either unaware of developments in assessment technology or happy to relegate consideration of them to some indefinite future date.

(b) Where a subject-related professional body existed (chemistry and medicine) it was seen as a very important change factor. Such bodies had not in fact prescribed particular changes in assessment (although the GMC had come close), but any changes proposed at local level had to be approved by such bodies.

(b) Chemistry, philosophy and design attributed some importance to either increased numbers of students, to increased diversity within the student population,

or to both. Medicine had seen no increase in the number of students and, being in high demand, rigorous selection had meant no obvious change in the range of abilities of students.

(c) Chemistry and philosophy, the traditional academic 'pure' subjects, ascribed some importance to a concern to develop/assess transferable skills; medicine and design did not. Medicine and design had both shown increased concern for the assessment of communication skills, but they saw these as important within the destination-professions, rather than ones which could be transferred to some other occupation.

(d) This researcher found it remarkable that chemistry and philosophy staff attributed so little importance to the influence of enthusiastic colleagues or of any educational/assessment theory or theorist. The position in medicine and design was distinctly different. The people interviewed there gave the impression of being more enthusiastic about assessment and more likely to be influenced by the enthusiasm of others. Medicine has a strong tradition of medical education and indeed of assessment research, but design does not. The interview data suggested an interest in, and openness to, educational influences in the two professional preparation areas that did not exist in the two academic subjects. It is tempting to argue that the concern for professional preparation may actually have encouraged awareness of the demands of educational professionalism. This did not mean that generic educationists were viewed with enthusiasm.

Future changes

Participants were asked to describe changes in assessment they personally saw as desirable and to speculate about changes that might be required of them. Detailed answers are in Chapters 5 to 8 and related appendices.

A summary of changes desired is given in Table 3. Again there were distinct disciplinary differences. For example, chemistry staff wished a clearer definition of the core content to be assessed, a wish entirely in line with its earlier location as a 'hard' subject; philosophy showed no similar concern. Design staff wished to see greater consistency in standards across internal specialisms; this was not generally specified by medicine as a desired change, but it was seen as desirable by those in charge of years 4 and 5 clinical specialisms. However, rather than dwell on the apparent differences, there are underlying themes which point up similarities, each appearing in at least three subject areas.

- (i) Students should be given a better idea 'of where they are' through more and more effective formative and informal assessment. This can be seen as a wish to give the learning-enhancement function of assessment a higher profile.
- (ii) The assessment of practical, clinical and communicative competence should increase, but assessment criteria should be clearer, more transparent and better understood. This seemed a desire for authenticity in assessment, but in a form which did not sacrifice reliability.

Chemistry	Philosophy	Medicine	Design
Define and assess core knowledge	Less dependence on essays and examinations	Give students a better idea of where they are	Comparability of standards across specialisms through shared criteria
Progress with objective and computer-based assessment	More formative and informal assessment	More formative assessment	More innovative methods
More reliable assessment of laboratory work	Reduced reliance on teaching assistants	Increased reliability of clinical assessment	More integrated assessment
Greater clarity and shared understanding of criteria	Greater clarity and shared understanding of criteria	More review of overall coherence	Review of overall coherence
Better assessment of transferable skills	More weight on, and consistency in, assessment of talk	Introduce a 'progress' test	

Table 3 Future assessment changes seen as desirable

(iii) There should be a move to more coherent and integrating systems of assessment. There was a wish in all four subject areas, made wholly explicit only in medicine and design, for a change in emphasis from looking at separate assessment episodes/methods to reviewing the nature and effects of assessment systems as a whole.

This last theme relates to the issue of communication about assessment, to which we return later. Interviewees wished the following to be reviewed and discussed in appropriate fora: the assessment of core knowledge in chemistry, over-reliance on graduate teaching assistants in philosophy, the potential of computer-based assessment in chemistry and medicine, more shared understanding in all subjects of the assessment criteria to be applied.

Table 4 summarises the changes that interviewees predicted might be required of them, whether they saw these as positive or negative, and what they said the source would be.

Again there were distinct disciplinary differences. To give just one example, threats to the validity of assessment from student dishonesty were a worry in both academic subjects, but in the professional-preparation areas this concern did not appear, perhaps because there was much more assessment based on the observation of student behaviour in practical settings.

Chemistry	Philosophy	Medicine	Design
More action to prevent plagiarism +ve <i>The University</i>	More action to preserve integrity of assessment +ve <i>The University or faculty</i>	Changes decided by internal monitoring +ve <i>Faculty</i>	
National standardisation of both content and assessment -ve <i>National policy</i>	Even more weight on continuous assessment -ve <i>The University</i>	Adoption of uniform grading system -ve <i>The University</i>	Uniformity in practice across specialisms -ve <i>Glasgow School of Art or the University</i>
More paper, more audit, more Bureaucracy -ve <i>National and University</i>	More weight on assessing transferable skills +/-ve <i>National</i>	De-coupling of graduation and registration +/-ve <i>Government and GMC</i>	More numerical and quantitative assessment -ve <i>GSA or University</i>
	Economy through computer-based assessment -ve <i>The University</i>	Economy through cooperation and computer-based -ve <i>National and University</i>	Economy through less dialogue with students -ve <i>National and University</i>

Table 4 Assessment changes, positive and negative, that might be required and sources of the requirement

There were three changes predicted from all areas. The first was a move to greater dictation of how assessment should be conducted in the interests of standardisation and uniformity across the institution; the second was the giving of even more weight to assessment documentation and the recording of evidence; the third was an imposed shift to more objective, quantitative, computer-based assessment in the interests of economy. These predicted changes were viewed negatively.

The sources of required future changes that were revealed in the unstructured part of the interviews were confirmed in the structured episode. Interviewees were asked to identify change factors they thought would become more influential in future. Responses are summarised in Table 5.

Chemistry	Philosophy	Medicine	Design
National policy	National policy	National policy	National policy
Institutional policy	Institutional policy	Institutional policy	Institutional policy
Faculty policy	Faculty policy		
Subject professional body (RSC)		Subject professional body (GMC)	
	Student pressure	Student pressure	Student pressure
			Student numbers
Assessment technology	Assessment technology		Assessment technology
	Assessment elsewhere		

Table 5 Factors influencing assessment change predicted to become more important

It is obvious that 'policy' was expected to become a much more influential factor. It may be helpful to remind ourselves what sources the 'policy' entries indicate and the nature of the influence policy was expected to exert. For the departments of chemistry and philosophy, 'faculty' refers to the Faculty of Science and the Faculty of Arts respectively. Medicine is itself a faculty; for design, the School of Design performs the functions of a faculty. For chemistry, philosophy and medicine, 'institutional' refers to policy perceived as coming from the University of Glasgow; the University may be the channel for supra-institutional policy and the Faculty may be the channel for University policy. For design, 'institutional policy' means policy coming from, or through, either the University or Glasgow School of Art. For all subject areas, 'national policy' refers to policy from the government and its agencies, particularly the Quality Assurance Agency. For chemistry and medicine there are relevant subject-based professional bodies which also act as sources of supra-institutional policy.

Future policy influence was expected to extend its range, from the more bureaucratic to the more educational aspects of assessment and to alter its nature, becoming more prescriptive and directive and less open to local interpretation and indeed resistance. Influences which did not mention policy were also interesting. It was felt in three subject areas that assessment might have to change in future because of an external 'pressure', from the general climate of opinion, to admit students to a more active partnership in assessment processes. Such a pressure might of course become an element in policy designed to sensitise people to such a priority. There was also a pressure perceived in rapidly developing computer technology; one interviewee observed that if academic staff did not respond

creatively to this pressure on their own initiative, then policy might be put in place to force them to respond. Assessment technology adopted for sound educational reasons might well be productive, but being forced to deploy technology because of a policy imperative based on the hope of cheapness was anathema.

The perceived origin of policy requirements had a significant influence on how favourably future policy directives would be received; policy from the QAA would be seen as unwarranted interference ('they don't understand the particular character of our subject'), whereas the same policy from a subject-related body would be seen as legitimate intervention. The strength of disciplinary culture was again evident.

SECTION 3: RELATING ASSESSMENT POLICY AND PRACTICE.

This study examined assessment policy and assessment practice in the University of Glasgow between 1998 and 2002. This section looks at the relationship between the two.

Chapter 4 addressed two research questions. The first of these was, 'Did the University generate new assessment policy?' In summary form, answers were as follows.

- Agreement was reached on a set of underpinning principles of assessment; these were sufficiently general to allow widely differing interpretations in different faculties and departments.
- An assessment Code was produced and incorporated within the formal University regulations; that code was incomplete; in June 2002 a crucial section of the Code was still being debated.
- A guide to good practice in assessment was promised, but not delivered.

- Policy activity was stimulated; assessment became a ‘hotter’ topic and rose higher on the policy agenda.
- There was a significant change within the University in the committee structure through which assessment policy was to be further developed.

Whether or not the University can be judged to have generated new policy on assessment depends on the meaning attached to ‘policy’. If this is construed as policy-making (in the sense of issuing products for acceptance and implementation), then the exercise largely failed. If it is construed as policy-activity (in the sense of a process which in the longer term might have significant effects on practice), then there was a measure of success.

The second research question was, ‘Did the new policy aim to change assessment practice?’ Again the answer was not simple; the general message can be summarised in four statements.

- New policy was intended to make unacceptable aspects of practice less likely.
- New policy products focused on the bureaucratic and administrative aspects of assessment; disagreement on aspects of a more directly educational nature was not resolved.
- There was a marked shift in intention, from the production of policy which emphasised University-wide conformity in assessment practice, to policy which required general principles to be adopted in ways appropriate within different subject-related contexts.
- There was acceptance amongst University-level policy-makers that the trend towards ever-increasing central direction of assessment practice was inevitable, given the apparent direction of supra-institutional policy influences.

Section 2 of this chapter provided the material for an answer to research question 6: 'To what extent was policy perceived as influential in bringing about change?' The question was concerned with both perceived influence on past changes and predicted influence on future change. Important points were as follows.

- Interviewees in all four areas infrequently used the word 'policy' in describing factors important in the past.
- Assessment changes of one type, 'policy-related shifts', were clearly seen as related to policy requirements coming from a source external to the department or faculty.
- General evolutionary trends in assessment were said to have happened in conformity with fashion or 'the spirit of the times'; this could be interpreted as an indirect policy influence, in that policy was likely to have prioritised attention to some aspect of assessment practice.
- Assessment change categorised as 'new-course introductions' was promoted by a policy requirement that course planning procedures include a re-thinking of assessment practice.
- Interviewees used the word 'policy' much more frequently when predicting what would be influential in future change in assessment.
- Future assessment policy was expected to be more wide-ranging, extending more into educational aspects of assessment, and more prescriptive.
- Interviewees distinguished between external policy from the University (or from government or QAA, channelled through the University) and policy from

subject-related professional bodies. The latter was viewed more positively than the former.

In chapter 2, 'early notions' of the likely relationship between assessment policy and practice were introduced. These must now be reviewed. The first of these was that 'assessment policy is an important cause of change in assessment practice'. Reviewing the notes taken at the early meetings of the Assessment Working Group, there is no doubt (a) that the group saw their task as to generate policy which the Education Committee and the Senate would endorse for implementation across the University and (b) that the purpose of the policy was to change assessment practice for the better.

The AWG assumed that the principles they identified and the courses of action they proposed would be adopted and that this adoption would bring about an enhancement of existing assessment practice. Although the members showed no sign of overbearing self-importance, they did believe they were engaged on a task that was important; as busy people they would not have devoted so much time so conscientiously over a lengthy period if they had not. In other words they assumed that the products from the group would make an important contribution to the outcome they desired i.e. to better assessment.

Whether or not the 'policy' they produced has actually caused significant changes in assessment practice is a different question. At present, it would seem that a great deal of policy effort has resulted in a flurry of policy-activity across the University, but remarkably little change in how students are assessed. The future impact of the 'policy' is of course unknown and it may be that detectable 'real' change in a very large and complex institution can only be expected to happen over

a protracted period. It is, however, difficult to avoid the conclusion that the AWG spent an inordinate amount of time discussing what the assessment changes should be and very little considering how change could be creatively encouraged and its introduction effectively managed.

From the perspective of the policy-generators then the relationship statement has to be modified: 'Although assessment policy has the potential to be an important cause of change in assessment practice, whether or not it actually does so depends on a wide range of other factors'. These factors include, but are probably not limited to, the following.

- (a) **Actual change may not add up to 'real' enhancement of practice.** Policy-makers may have the organisational power to require assessment change, but the changes which actually occur will not add up to educational progress unless practitioners perceive the changes as beneficial when judged against their own values and their other commitments within their subject-related territories.
- (b) **Increased effectiveness of existing practice is easier to achieve than fundamental changes in the methods and procedures of practice.** Policy-makers find it easier to bring about first-order changes relating to the efficiency and effectiveness of what is currently done than second-order changes involving new approaches to, and methods within, assessment practice (which inevitably involve changes in educational assumptions and theories).
- (c) **Responses to policy depend on shared understandings.** Individual and collective responses to proposed policy depend on the existence of shared and agreed understandings of what the policy is intended to achieve; such understanding requires what Fullan called interactive professionalism. Without the conditions for,

and encouragement of, such interaction, policy impetus is dissipated and policy implementation distorted.

(d) Responses to policy depend both on its implications for work and on how people feel about it. Policy-makers should try to take into account that the response to new policy will depend both on how the policy affects what people do and how they feel about what is required of them. It will be helpful if they understand that responses can range from acceptance and implementation to reconstruction, avoidance and demand-reduction. The response to policy-promulgation may also be an increased involvement in reactive activity, either in resistance to the policy or to alter the conditions of practice.

(e) Policy should be about both institutional and individual renewal. Educational leadership may well have an obligation to exert policy pressure from above to bring about assessment change; equally it has an obligation to support individual and local departmental initiatives for change. Re-visiting the documentation made available to, and emanating from, the assessment policy group, it strikes this researcher as remarkable that the emphasis was always on policy exerting pressure from the centre on the periphery to change; on no occasion was there any explicit mention of providing support to individuals in their local attempts to innovate.

Turning now to the practitioners' perspective, the relationship statement again needs to be modified. Assessment policy was seen as one cause amongst others of past changes in practice; its direct importance was perceived as relatively low, but the indirect influence of policy was more significant (a) through the influence it had on the assessment zeitgeist and (b) on its prescription that

assessment must be a component of course-planning. It was clearly predicted that assessment policy would become a much more important cause of change in the future. 'Assessment policy is an important cause of change in assessment practice' should thus be re-stated as 'Assessment policy is expected to become a more important cause of change in assessment practice in the future than it has been in the past'. The explanation for this is straightforward; there had been an increasing amount of central direction within the University and an increasing amount of direction of the University from external policy sources; both trends were expected to continue. (It should be remembered that interviews were carried out in session 2000-2001; the alteration in the discourse used by QAA from precepts to guidance did not occur until April/May 2002.)

We have so far been dealing with the relationship of policy to practice, based on an assumption of temporal precedence of the former i.e. policy comes first. 'Early notion' 2 (Page 40) was that 'Existing assessment practice causes the emergence of new policy'. It was noted that this relationship could apply at national level: 'Assessment practice in our universities is very varied in nature and quality; there is thus a need for policy aimed either at regulation or quality enhancement development'. Equally well it could apply at institutional and faculty/departmental levels. Existing assessment practice may be such that those with the authority to make policy may decide that new policy is required to bring about change.

There was clear evidence of this relationship applying at institutional level in this study. There were examples of assessment practice in some areas which central management of the University judged to be quite unacceptable. This was given by the Vice-Principal as an important reason for setting up the AWG. If there was an

aspect of assessment practice which was deemed inappropriate in relation to existing policy, then no doubt a head of department could be quietly told to change the practice. An authority figure was able to back up any directive by reference to the policy. Where the existing policy was outdated or insufficiently comprehensive, then there was a strong justification for developing new policy.

These examples seem to place practice in a deficit frame i.e. the practice was in some way inadequate and policy change was then a mechanism for pulling 'defective' practice into line. This probably stemmed from a perception that national policy frames the matter similarly. There is evidence, however, of recent attempts to change the policy emphasis from regulation of the deficient to the development of all. (See e.g. SHEFC, 2002) The current orthodoxy now promoted is that the question is no longer 'Is assessment practice good enough?' but 'How can the quality of assessment practice be enhanced?' The underpinning stance has shifted from 'some places are not good enough' to 'nowhere is so good that it could not be better'. Even this shift, however, still carries an implication that practice-improvement should be policy-led. The converse should be recognised: policy-improvement can be practice-driven. An obvious example might be some innovative on-line course which made comprehensive and appropriate use of computer-based assessment. It can easily be envisaged that existing policy would not adequately cover procedures within such a course. In other words, good assessment practice points up a policy deficit. There was no similar example within this study; the nearest parallel came from the new medical curriculum. Here a new assessment system had been devised; a new assessment policy was contemporaneously emerging within the institution and the view of the medical faculty was that their carefully thought-out assessment system would be seriously compromised by the

implementation of the University's proposed grading system. Intense political activity resulted in a significant change in the new policy in the direction of 'where there is specific need, there can be specific provision'. The medical faculty could be assumed to share the institution's unifying purpose without subscribing to any policy promoting identity of procedures. Practice had thus prompted the pulling of policy into line.

There has so far appeared no powerful reason to reject the original statement ('existing assessment practice causes the emergence of new policy'); it could appropriately be expanded to 'existing assessment practice is one factor which may cause the emergence of new policy'. Could there be other reasons for policy generation which might even be more potent? This prompts a more careful examination of the original statement. It is not practice as such which reveals some need for policy development, but rather the perceptions of people that there is some mismatch between what assessment practice is and what it ought to be. The wording of the 'early notion' focuses on assessment practice as it is; there could equally well be a statement which emphasises that changing views on what assessment ought to be like can cause the emergence of new policy. The Australian example is helpful here: a national policy was issued to proscribe assessment systems based on norm-referencing. One could construct an explanation that norm-referencing practice caused the change in policy; or alternatively that policy-makers thought that assessment systems ought to be criterion-referenced; or better still that the new policy came about from the perception of policy-makers that a mismatch existed between current practice and what assessment practice ought to be like. The statement of the 'early notion' could be re-worded to take this into account: 'A

mismatch between existing assessment practice and what policy-makers think it ought to be, causes the emergence of new policy’.

SECTION 4: ISSUES EMERGING AND IMPORTED

Meanings attached to ‘issue’ were outlined in Chapter 2 and Appendix III.

Emerging issues

The issues which emerged in the four subject-areas are described and discussed in Chapters 5 to 8 and related appendices. They are summarised in Table 6.

Chemistry	Philosophy	Medicine	Design
<p>The burden of assessment</p> <ul style="list-style-type: none"> - wearisome workload - intolerable at peak times - expected to increase 	<p>Lack of originality</p> <ul style="list-style-type: none"> - in student responses and in questions set - decline in 'eccentricity' or increase in professionalism 	<p>Appropriate grading scale</p> <ul style="list-style-type: none"> - different scales needed for understanding and for competence - resistance to University policy on uniformity 	<p>Valid assessment of creativity</p> <ul style="list-style-type: none"> - what counts as creative - assessor consistency - protection of students from injustice - local or global validity of creativity judgements
<p>Plagiarism</p> <ul style="list-style-type: none"> - increasing student dishonesty in assignments, reports, group work - plagiarism seen as barrier to innovation 	<p>Integrity</p> <ul style="list-style-type: none"> - increasing student dishonesty except in examinations - concern over inconsistency and bias in assessors 	<p>Workload shifts</p> <ul style="list-style-type: none"> - different people now carry burden of assessment - problems of authority and communication 	<p>Appropriateness of honours degrees</p> <ul style="list-style-type: none"> - irrelevance in design - negative effects on students
<p>Objective and computer-based assessment</p> <ul style="list-style-type: none"> - distrust of selected response methods - need for methods economical of staff time - lack of knowledge of methods and their potential 	<p>Drift in standards</p> <ul style="list-style-type: none"> - response by assessors to expectations of 'good' degree - encouraged by diversity in assessment methods 	<p>The expertise of assessors</p> <ul style="list-style-type: none"> - new methods required better informed thinking - not all assessors knew or cared enough 	
<p>Examination stress in students</p> <ul style="list-style-type: none"> - opinion divided on need to replace stressful terminal exams 	<p>Staff workload</p> <ul style="list-style-type: none"> - the time to do assessment means no time to think about changing it 	<p>Authenticity and reliability</p> <ul style="list-style-type: none"> - worry about how to increase consistency without losing reality connections 	

Table 6 Issues emerging in the different subject areas

In each column the issues are given in decreasing order of salience. The amount of researcher judgement involved in selecting issues and in imposing an order of importance should be recognised. Even the number of issues is somewhat arbitrary. There appear to be four issues in each of the first three subjects, but a case could be made for their being only three in philosophy. (Was 'standards drift' genuinely an issue or only a matter of mild disquiet?) Perhaps there should be five in medicine. (The need for an assessment review mechanism could be treated as an emergent issue, rather than being dealt with under the imported issue of communication.) In design, how creativity could be validly assessed was a very big issue, about which all participants agonised at length; there may have been other issues (in addition to the place of honours degrees within design) that would have emerged given time. The degree of salience was a researcher judgement based on two criteria, the length of time devoted to talking about the matter and the strength of concern being voiced.

It is worthy of comment that issues emerged at all. They were not prompted by any special technique; all interviewees were busy people who could have reduced the length of the interviews by providing only brief, focused answers to the questions posed. That they devoted time to introducing and then amplifying their worries and concerns stemmed from a recognition of the inherently problematic character of assessment and indicated the seriousness with which they were prepared (in the interviews) to reflect on their own assessment practice and that of their colleagues. On listening again to a selection of the interview tapes this researcher was struck, not only by the laughter of shared humanity, but by the thoughtful engagement of minds. Disconcerting was the number of times people

said things like ‘I didn’t think I had so much to say about assessment’ and ‘No-one has ever listened to me so attentively on assessment before’.

Examination of Table 6 prompts the reaction that there were clear subject-related differences in assessment thinking. Only the chemists were concerned about objective and computer-based assessment; only the philosophers identified a downward drift in academic standards as an issue; only the medical staff were concerned about changes in the distribution of the work of assessment; only the design staff talked at length (the philosophers spoke briefly) about the problem of assessing creativity. While there were undoubtedly disciplinary differences, there were also some similarities and common themes underlying some issues appearing in two or more columns.

1. In all four areas there was a detectable concern within the issues for what was earlier christened **epistemological alignment** i.e. that assessment practices should be congruent with the perceived nature of the subject taught. The chemists felt that they should be making more use of objective and computer-based assessment but they were worried that their subject would then be distorted by the selected-response items they saw as inevitable with objective testing. The philosophers felt that the questioning character of their subject required assessment procedures which encouraged students ‘to risk going out on a limb’ and had serious misgivings about how inappropriate assessment encouraged reproduction by students of lecturer-delivered orthodoxy. The medical staff feared that university policy on grading was over-influenced by those non-medical academics who did not appreciate the nature of medicine. In both medicine and design ‘the nature of our subject’ was extended to cover the nature of the work done by qualified

practitioners i.e. epistemological alignment shaded into occupational authenticity.

As students progressed through their medical course it was regarded as essential that assessment came closer and closer into line with the medical practice into which students were being inducted. In design, all assessment criteria related to a definition of the characteristics of the design process and the main assessment method (portfolio assessment) was justified by the nature of design practice.

2. In the issues in all four areas there was a concern for integrity in assessment. Validity ('the extent to which an assessment actually does what it claims to do') may be an adequate concept to cover this concern, but the word integrity has been chosen to emphasise that not only have assessment procedures to be planned with validity in mind, but also that all reasonable steps have to be taken to minimise threats to validity from whatever source.

Student dishonesty was clearly identified as a worrying threat to validity in both the 'pure' academic subjects, chemistry and philosophy. Dishonesty was believed to be more likely to occur with essays, assignments and project reports and less likely in traditional examinations. The fear of growing dishonesty amongst students inhibited innovation and could result in pressure mounting against any further diversification of methods – a 'regression' to conventional assessment under examination conditions. To some extent this fear of dishonesty was related to the number of students involved; both chemistry and philosophy dealt with very large classes. On its own this is an inadequate explanation; medicine deals with equally large numbers and for the medical staff student dishonesty was not an issue. The crucial factor seemed to be the extent to which students were assessed by staff who did not know them and what they are capable of; this lack of knowledge of students could allow dishonesty to go undetected. On the other hand it was argued that

knowing students well increased the risk of assessment being biased by irrelevant personal knowledge.

The view that assessment by examination does not give much scope for student dishonesty may well be sound, but the converse (that non-traditional assessment methods are riskier) requires more scrutiny. The assessment of practical work in a laboratory is not vulnerable to plagiarism, although assessment of laboratory reports may be. The valid assessment of skills in oral presentation is not threatened by personation, although clearly there could be some dishonesty in acknowledging sources. The assessment of clinical competence in-situ and of artefact production in the design studio offer little scope for dishonesty; neither do oral and viva-voce examinations.

Assessor inconsistency and bias were also recognised as potential threats to the validity/integrity of assessment. There had been considerable doubts about the fairness of the assessment of clinical capability within the old medical curriculum; too many clinical assessors were known to have particular hobby-horses and alleged to possess gender and/or race bias. Clinical examinations were much changed in the interests of consistency and fairness, but interviewees were still worried about the inherent unreliability of such assessment. In design, classes were relatively small and all assessing tutors knew well the studio activities of their students and assessment dialogue was carefully developed over the years of the course. The potential here for assessor bias was clearly recognised; there was great emphasis within the interviews on the procedures in place to protect students from possible injustice. The researcher was almost convinced that design assessment had much more integrity than existed in the other areas. This could be so, but there is one caveat. The consensus judgement of many assessors would not possess integrity, if

that consensus was over-influenced by individual assessors exercising power rather than applying agreed criteria. Some design interviewees suggested that their assessment operations had their murky moments.

3. In both the 'applied' areas there was concern about **the appropriateness of grading systems**. In these areas there was a reaction against the spreading out of students over a range of grades as happened when using A to E scales and Honours degree classification. There was a distinct preference, in the areas of clinical performance and studio practice, for coming only to the binary decision 'Yes, you have satisfied our criteria of fitness-to-practise' or 'No, you are not yet satisfactory'. This was somewhat complicated in the majority of interviewees by their wish to identify the most meritorious students with some kind of label that indicated their excellence.

The appropriateness of the grading scales currently used, and of others being proposed in new policy, did not appear as an issue at all in the 'pure' subject areas. The value of a traditional norm-referenced spread of student grades was rarely questioned in these two subjects.

4. In three areas, some aspect of **assessment workload** was an issue. In chemistry and philosophy, people said that although the burden of assessment work was not yet intolerable, it might well become so. There was some disagreement about what increasing workload implied for innovation in assessment. Perhaps changes should be sought which promised economies in staff time without any unacceptable decline in assessment quality; some chemists saw hope in computer-based assessment. On the other hand, while time could be found for carrying out

assessment within the present procedures, time could **not** be found to consider innovations with the care necessary to avoid negative effects; more than one philosopher said things like, ‘We have time to do assessment, but not to think about changing it’.

The situation was different in medicine; there was a distinct impression here that the problem of staff assessment overload had not yet emerged, perhaps because of the motivating and anaesthetising excitement of implementing a radically new curriculum, under the interested gaze of many spectators. What had emerged was the realisation that different assessment methods had resulted in quite radical shifts in who did the work – different people were now shouldering the burdens. The issue that emerged was one of the **management** of workload shifts.

In ending this section, attention is drawn to an issue in the table only identified explicitly by some chemists. (It was an issue also for one of the medical staff.) This was the possible damage done to students by stressful, ‘big-bang’ terminal examinations, damage in the sense of having a seriously adverse effect both on the quality of performance displayed at the time and also on longer-term attitudes to learning. These chemists saw this as a problem which merited serious changes in assessment in future; the medical person saw it as a problem which had been serious in the old curriculum but had, at least partly, been resolved in the new assessment arrangements.

Imported issues

Data from the interviews were analysed with reference to five 'researcher' issues. These were chosen during the design phase as ones which assessment 'experts' would agree were of topical interest and also as ones which the researcher could argue were linked in some way with the general concept of professionalism in assessment. What interviewees said about these imported issues is presented in summary form in Tables 7 (a) and (b). Where there are numbers in brackets, these indicate the number of people holding that view.

	Chemistry N=8	Philosophy N=8	Medicine N=11	Design N=9
Dominant assessment purpose	Learning enhancement should dominate (5) Certification and accountability (3)	Learning enhancement should dominate (8) Certification is inescapable (8)	Certification inescapable (11) Explicit emphasis on accountability (2) Learning enhancement equally important (11)	Accountability to public (7) Certification inescapable (9) Learning enhancement equally important (9)
Constructive alignment	Constructive alignment mentioned (1) Content alignment only (1) Nature of subject (7)	Constructive alignment mentioned (2) Nature of subject (8)	Constructive alignment mentioned (8) Nature of medical practice (10)	Constructive alignment mentioned (2) Nature of design practice (9)
Communicating about assessment	E-mail consultation Some in committees and corridors Less than in past - not enough	E-mail consultation About individuals - not about issues Less than in past - not enough	Effective on some new methods Need for discussion forum Overall, less than ideal	Whole system built on dialogue

Table 7 (a) Interviewees' views on imported issues

(Numbers indicate numbers of people holding that view.)

	Chemistry	Philosophy	Medicine	Design
Criterion-referencing	Norm referenced (3) Grappling with transition (5)	No clear diagnosis - probably norm-referenced	All grappling with mixed system Norm-referencing for administrative ease Criterion-referencing an educational obligation	All happier with criterion-referencing Norm-referencing a nuisance from admin
Constrained autonomy	Acceptance of current constraints Prediction of more authoritarian constraint - bureaucratic and educational	Acceptance of current constraints More external direction predicted	More constraint expected - acceptable from GMC - to be resisted if from University	Little external constraint at present More expected from University and QAA

Table 7 (b) Interviewees' views on imported issues

1. Dominant purposes of assessment

Assessment can serve many purposes. It is common to conflate these into three broad purposes, usually labelled certification, accountability and learning enhancement. When the first two are combined, they constitute the control function of assessment; the third is the growth function. (Nisbet, 1993; Broadfoot, 1996)

There was an assumption that what interviewees had to say about the broad purposes of assessment would allow inferences to be drawn about the conception of professionalism that they possessed. There was an expectation that participants would see certification/accountability purposes as inevitable and would then go beyond these to reveal some commitment to ensuring that their assessment practice promoted growth in student learning i.e. the learning enhancement purpose. Inherent in these expectations was the view that academic educators must show an informed concern for the public interest: only those students who reached the required standards should be assessed as successful (i.e. certification) and assessment should provide evidence that courses were being effectively taught and provided value for money (i.e. accountability). Such concern can be taken as a necessary but not sufficient condition for education-professionalism. Higher education teachers should be aware of the role of assessment in serving the interests of their primary 'clients' i.e. their students.

The majority of the chemists and all the philosophers professed their view that learning enhancement should be the dominant purpose of assessment. A minority of the chemists talked only about preserving standards in those to be 'certificated'. All of the medical and design staff talked about the inescapable purpose of assessment for them, that it should determine whether students were 'safe-to-practise' or 'fit-to-practise'. Few medical staff placed any obvious

emphasis on the accountability purpose; most design staff felt that 'going public' in shows and displays showed their commitment to accountability. All medical and design interviewees went on to talk about assessment having an equally important role in encouraging students to learn more, to learn more effectively and to learn more of the 'right' things. In general then, what interviewees said on this issue provided evidence of their professionalism as assessors.

2. Constructive alignment

Constructive alignment is a concept that has only appeared in the assessment policy literature and in academic texts in the last five years or so. (Biggs, 1999a, 1999b) It refers to bringing statements of intended learning outcomes, teaching methods and assessment into a clear and productive relationship. It was chosen as an imported issue because talk about it by interviewees could act as a useful indicator (a) of their understanding of how assessment could have a positive impact on the quality of student learning and (b) of their awareness of recent assessment policy and discourse. Professionalism in assessment, it is argued, requires a knowledge base, indeed a measure of assessment scholarship; such scholarship would be somewhat inadequate without any understanding of constructive alignment. Possession of the concept might of course be inferred to exist without there being explicit use of the precise label.

In each of chemistry, philosophy and design one or two people explicitly used the term; these people had all been involved in quality assurance committee work at faculty or university level. Their knowledge of the term derived from their knowledge of the policy literature; one person in design referred to academic assessment literature, his reading stimulated by collaboration in assessment

planning with European colleagues. The position in medicine was remarkably different; the majority of those interviewed either used the term or clearly showed their familiarity with the concept. Policy-led procedures for planning the new curriculum required the idea to be taken on board by those responsible for assessment; what began as a new-fangled bit of jargon became an accepted and unremarkable bit of specialised terminology.

The idea that assessment had to be productively aligned with **something** other than intended learning outcomes occurred in almost all the interviews. One chemist talked exclusively of assessment being in line with content coverage. Most people in chemistry and philosophy expressed a belief that their assessment methods could be defended by epistemological alignment (see Page 215). In **both** chemistry and philosophy most assessment still relied on conventional written examinations; it is difficult to see how this can be defended by any epistemological alignment argument. In medicine and design, support for their assessment practice came from ‘occupational alignment’.

The general conclusion is that most medical staff were more aware of, and possessed more understanding of, this idea in contemporary assessment discourse than did people in other areas. The challenge of planning and implementing a radically new curriculum had encouraged medical staff to update their assessment knowledge base, to acquire more assessment scholarship. Possibly more were made aware of the need to do this than actually did it; it will be recalled that one of the emerging issues in medicine was that assessment expertise was said to have developed unevenly.

3. Communicating about assessment

Communicating about assessment, face-to-face in a collegial context, seems to this writer essential both to develop and to sustain professionalism in assessment. There are four strands of argument leading to this conviction. The first concerns reliability in assessment. Reliability begins with the formulation of assessment criteria. Mere existence of criteria is not enough; criteria need exemplification for meaning to be attached to them. But this is still insufficient; shared understanding of the criteria only happens when assessors form part of a regularly inter-communicating group. (See Wolf, 1995, p 77) Secondly, it can no longer be assumed that boards of examiners/assessors communicate much about the aggregation of component grades and decisions about course awards and degree class, let alone about the validity and reliability of the methods by which student learning is assessed. Decision-making has increasingly become mechanistic and formula-driven. Thirdly, although one can imagine a group of assessors grown familiar with each other over time maintaining the assessment status quo without much communication, it is very hard to conceive of significant change in assessment being planned and implemented without intense periods of discussion and communication. Fourthly, hard-pressed academics have got used to communicating about many matters, including assessment, through e-mail interactions; this mode of communication appears ill-suited to debate about difficult principles of assessment.

A group of the chemists reported conversing about controversial aspects of assessment in coffee-room and corridor; a different group claimed discussion of assessment within the Teaching and Learning committee; it was also said that these two groups did not connect with each other. In philosophy there was plenty of

communication about disagreements in the assessment of individual essays and dissertations and about decisions on border-line students; the 'bigger issues' in assessment were not discussed in meetings of boards of examiners or in routine departmental meetings; people put any concerns 'on the back burner'. In medicine, face-to-face discussion was built into some assessment procedures (MEQs and the MILE) but there was a strong view that there was no appropriate forum, or time, to discuss the overall coherence of assessment. In all three subjects there was agreement that there was less inter-personal communication than in the past and that currently there was too little. The position was very different in design; assessment was incontrovertibly a matter of dialogue, both between assessors and students and amongst assessors. The position was regarded as healthy and not much different from the past; communication was said still to be frequent in both informal and formal settings.

If it is a requirement of professionalism that there be effective communication amongst the assessment partners, then the overall picture was not reassuring. In three areas (chemistry, medicine and design) there had been intensive discussion to plan the assessment required in new courses, but it was questioned whether adequate communication was continuing to monitor and support the changes introduced. It was noted on Page 36 that Fullan saw 'interactive professionalism' as a prerequisite of 'real' change; in this research the data suggested that interactive professionalism was generated by the requirements of course planning but did not stay in place to ensure that innovations were sustained to best effect. Where there were no new qualifications or curricula, there was inadequate communication about problematic matters and about whether within-course innovations were desirable. Such a claim seems valid for chemistry and

philosophy; the position for design was less clear. Either continuing dialogue supported an assessment scheme which justified the general approval it was accorded or communication was not sufficiently regular or deep to prevent a proper confidence edging into complacency. Assessment in design was ‘good enough’; should the assessors have been doing more to ensure they stayed ‘as far ahead of the game’ as they had been in the past? Overall, interviewees seemed to communicate more about their assessment worries with the researcher than they routinely did with their colleagues.

4. Criterion-referenced assessment

Criterion-referenced assessment was chosen (partly it has to be admitted because of an attack of compulsive alliteratititis – the CAs) as a short-hand indicator of the extent to which assessors had moved from unquestioning attachment to a traditional norm-referenced orientation to grappling with criterion-referencing or the ambiguities and tensions of mixed-mode assessment systems.

After the analysis was completed, the writer attended a lecture/seminar by Boud (2002) in which he outlined his view that there had been four developmental stages in assessment thinking: conventional assessment, educational measurement (which reached its high point in the 1960s and 1970s), competency and authentic assessment (increasing in prominence in the 1980s and 1990s) and sustainable assessment (his desired stage for the 2000s). Sustainable assessment was defined as ‘that assessment which both meets the needs of the present and prepares students to meet their own future learning needs’. This stage model provides a more useful framework than that indicated by the summary label ‘criterion-referencing’; the researcher should attempt to diagnose the extent to which interviewees had moved

beyond conventional assessment and educational measurement, whether they were involved in competency and authentic assessment, whether or not they had any vision of sustainable assessment. What interviewees had to say about criterion-referencing (or perhaps sustainable assessment) would be another indicator of their 'assessment scholarship' and of their commitment to the requirements of educational professionalism.

The thinking of three chemists had not moved beyond attempting to see everything in assessment from a norm-referenced perspective; the others were grappling, not always very coherently, with the tensions and ambiguities of mixed-paradigm thinking and procedures. The philosophers did not say enough on the issue for any confident claim about their stance; the few data there were indicated a largely unexamined norm-referenced orientation. The medical staff were all wrestling to reconcile norm-referenced assessment of academic understanding with criterion-referenced assessment of clinical competence. The majority saw criterion-referencing as an educational desideratum, but recognised that for some administrative purposes it was convenient if students were spread-out across a spectrum of achievement. The design staff were happy to operate a system in which criterion-referencing was dominant; they regarded the requirement to adapt to norm-referencing 'to please the people in the Registry' and to provide classified Honours Degrees, as an irritant.

What of the Boud developmental stages? In the documentation for the new medical curriculum, it was stated that by encouraging self-assessment capabilities in undergraduate students they would be helped to meet their own future learning needs. This hope seemed to carry through into practice only in the earlier years of the programme, in the self-monitoring built into the process of problem-based

learning. In the design curriculum, self-assessment was encouraged much more consistently throughout all years. Thus in the views from medical and design staff, there was evidence of understanding of the concept of sustainable assessment. Chemists and philosophers did not provide such evidence. It would appear then that philosophy staff were still at Boud Stage 1 (conventional assessment); most of the chemistry staff were at Stage 1, although some wished to move on to Stage 2 (educational measurement) and Stage 3 (competence and authenticity); medical and design staff were at Stage 3 with some residual traces of earlier stages and promising indications of moving on to Stage 4 (sustainable assessment).

If professionalism requires awareness of current thinking and assessment scholarship, then perhaps the most optimistic diagnosis that can be made is that staff in all subject areas would benefit from further professional development.

5. Constrained autonomy

Individual assessors can never be completely unconstrained; they have to acknowledge the legitimate constraints of departmental and other policies, of loyalty to the nature and traditions of their subject, of respect for academic standards and of commitment to ethical principles. Nevertheless, there remain difficult questions surrounding the extent to which groups of assessors within a subject should be free to decide how best to assess student learning in their subject area and also about the extent to which an individual assessor is to be trusted to take discretionary decisions. It has always been a characteristic of professionalism, that the professional has had the freedom to exercise professional judgement when precise rules and prescriptions are not available or appropriate. It was thus of interest in this

research to discover whether the participants felt that their autonomy within assessment had been, was, or might be unreasonably constrained.

There was little evidence that interviewees found existing constraints particularly irksome or unreasonable; there were objections to those policy requirements that seemed to put more emphasis on the documentation and evidencing of assessment than on its effective conduct. A few thought that contemporary 'audit culture' was itself a barrier to assessment innovation.

Conspicuous and striking, however, was the near consensus view and fear that in future their independence and autonomy would be further and undesirably constrained. External direction and central control were expected to become more oppressive and authoritarian as they extended from the bureaucratic to the educational and became less open to local interpretation, reconstruction and resistance.

Interviewees provided no evidence that policy had de-professionalised them; they feared that it might in the future. Whether that fear was justified remains to be seen.

SECTION 5: THE IMPACT OF POLICY AND PRACTICE ON PROFESSIONALISM.

The seventh research question was: 'What impact did changing assessment policy and practice have on professionalism?' This may look straightforward; it is not. The writer has found it so difficult to wrestle with that he has been on the point of abandoning it for some very different question. However, the conviction remains that the question is important and the overall structure of the thesis demands that it be addressed. Unpacking and clarification are necessary in preamble.

(a) The word 'impact' may be unhelpful. This is most obvious in relation to the effects of policy. 'Impact' suggest a physical metaphor from the snooker table. One ball is moving (assessment policy is changing) and knocks into another ball (professionalism): if the latter is static, it starts to move; if it is already moving, it moves in another direction. However, this kind of physical metaphor is of limited value in complex, social settings. It might be more fruitful to talk of the introduction of a revised script (policy) into an ongoing drama (defining professionalism and behaving professionally). The actors adopt the script, but implementation involves interpretation; the developing drama will not be exactly as envisaged by the author.

(b) Professionalism is a concept with an important place in the theoretical framework suggested by the present writer. However, it is also a concept that the participants in the research may or may not have or, rather, are likely to possess in different forms. Policy is an important feature of the context in which the research participants, as academic teachers, are required to operate. Changes in policy can clearly affect the conditions for the operationalising of professionalism; they can restrict professionalism or, in other circumstances, encourage it. There is a sense in which the concept of professionalism has some permanence, although the ways in which it is manifested are influenced by the particular context. Policy thus need not change what counts as professionalism itself.

On the other hand, policy can precipitate a re-think by academic staff of what the concept of professionalism means to them. In this case, changing assessment policy has an impact on the ways in which professionalism is conceptualised.

(c) There are also problems in relation to the impact of changing practice on professionalism. The very posing of the question in this way involves two separate reifications: a bundle of activities materialises into a thing called practice; a bundle of attitudes materialises into a separate thing called professionalism; the former is conceived as having an impact on the separated latter.

The assessment practice of an individual can be regarded as the behavioural component of a general attitude, rather than as something distinct from it. However, if assessment practice is construed as a generalised notion encompassing the collective activities of a subject-based group of assessors, then the expectation of conformity to the norms of group practice affects the ways in which individual members of the group operationalise their existing conceptions of professionalism. The individual may have to adjust what s/he counts as professional activity within the context of departmental/faculty activity. One can then conceive of changing assessment practice bringing about the suppression of some aspects of professionalism and the encouragement of others. In this sense changing practice can have an impact both on what counts as professionalism and on how it is translated into action.

(d) The question asked was, 'What impact did changing policy and practice have on professionalism?' There is an implication here that the research was only concerned with changes and their impact within a precise period (1998-2002).

An internal policy-generating body (the AWG) was studied during the specified period; the external policy context was examined from the date of the Dearing Report (NCIHE, 1997). In interviews with academic staff they were asked to talk about assessment changes they regarded as significant, irrespective of when these occurred. It would have been possible to restrict interviewees to changes

which had occurred since, say, September 1998, but it seemed unwise so to limit them and indeed to require them to remember when exactly something had occurred. Nor were they restricted to talking of policy influences that had emerged only since 1998; they had interesting things to say, for example, about the effects of earlier modularisation and course design policies. Just as earlier policy could have a continuing impact, so too would current policy have an impact well beyond the period studied.

Harsh practical realities have to enter into research design: limits have to be set. This is an argument for research of this type to be supplemented by historical study and by longitudinal studies of continuing impact.

(e) In the early stages of planning this research, it was decided that changing policy and practice would be the main foci of the empirical work of the research; they would be studied directly. Although professionalism was the third central concept it was decided **not** to make it an explicit focus of the research activity.

Why?

The researcher pondered at length whether to question research participants directly about their professionalism. It seemed unlikely that they would say much in response to questions like 'How do you currently conceptualise academic professionalism?' or 'How does emerging assessment policy influence your concept of professionalism and how it is operationalised?' He thus decided that if people were encouraged to talk in depth about personally significant change in practice and policy, they would probably say much about issues that concerned them and that from these data their stance on professionalism could be inferred.

Having taken this decision he continued to be troubled by doubts. Invited to provide a seminar/workshop on the research to University staff, he designed three

group work tasks organised around direct questions on policy, practice and their relationship to professionalism. (See Appendix XI)

The answers were not wholly without interest and tended to confirm some findings from the interviews. However, given the energy and time put into the exercise by a group of interested academic staff, the answers were strikingly thin. The outcomes of this exercise confirmed that direct questioning about professionalism was unlikely to be very productive. Some doubts were allayed, but the possibility remains that professionalism should have been an overt focus of the interviews but approached in a more imaginative way.

(f) The policy strand of the research involved some participant observation of a policy-generating group and interviews with key players; the practice strand involved interviews only. The obvious question arises, should there have been observation of practice, or indeed of professionalism in action? The only reason for its exclusion was severely practical: limits had to be set. This does mean that the research deals with what people said about assessment practice and what they said about issues that could be plausibly related to professionalism. It did not deal with what people actually did when assessing. This was not an observational study; the inevitable consequence is that whatever is said about 'impact' has to be a matter of high inference. There is a real risk of over-interpreting the data and of making claims of dubious validity.

If research question 7 is allowed to remain, it should be understood as a shorthand version of the following. 'From what interviewees said about changing policy and practice and about current issues in assessment, what plausible claims can be made about the influence of changing policy and practice on the conceptions

of professionalism possessed, on the ways these were operationalised and on how they might change?' Claims would draw on (i) definitions of professionalism given in Chapter 2, (ii) the changes in assessment that policy was intended to bring about, (iii) interviewee perceptions of significant past and future changes in practice and factors influential in bringing these about and (iv) emerging and imported issues. It would also deal with differences in disciplinary thinking and practice. The complexity of such a task is overwhelming. Suppose for each of four subject areas discussion were to be structured by a grid with seven key features of professionalism; the other axis would cover policy-change implications (say 2), practice implications (say 3), inferences from four emerging issues and the five imported issues. This would mean there should then be 392 interpretive essays followed by some synoptic treatise. The author actually embarked on this type of structured analysis, but the absurdity of the task provoked hysterical laughter. But, there must surely be some claims about 'impact' worth making. What follows is an attempt to identify these; the language of the headings can be related to the stipulative definition of professionalism given earlier.

a. Assessment was accepted as an implicit contractual obligation.

With three exceptions, everyone interviewed was a full-time permanent member of academic staff. (The exceptions, all in medicine, were two part-time academic staff and one member of 'academic-related' staff.) They all had contractual commitments to teaching, research and administration. An obligation to assess student learning was not explicit in these contracts, but all assumed that assessment was part of their job. When people were asked if they saw assessment as teaching or administration, the answer was always the same: 'Assessment is just

added on to your teaching, although some of the work is undoubtedly administration'. In some interviews, there was discussion of whether assessment should be seen as a separate activity from teaching and whether there might then be people employed as 'assessment specialists'. Again, all views were the same. One philosopher put it thus, 'It would be disastrous to separate assessment from teaching; assessment of students must remain the responsibility of those who teach them'.

There is a problem here. When a concern for assessment is not made explicit, it may be marginalised and not receive the attention it deserves. (The writer works for a 'Teaching and Learning Service' and regularly argues that it should be re-named the 'Teaching, Learning and Assessment Service' to help assessment get the emphasis, time and resources it merits. The name of the Institute for Learning and Teaching should be similarly expanded.) However, if assessment is separately identified, it seems to suggest that teaching and assessment are separate activities and on educational grounds perhaps it is better that they should not be seen thus.

One aspect of changing practice was that demonstrators and graduate teaching assistants were increasingly taking over some of the work of assessment. It was not always clear what aspects of assessment they could legitimately undertake and how they should be prepared and supported. Here was a group for whom contractual obligations could with benefit be made more explicit; there was an urgency here that did not, as yet, extend to full-time staff. The maintenance of professionalism in this changing aspect of assessment practice was thought to need the generation of new policy.

b. The seriousness of assessment was acknowledged, but not translated into time made available for it.

Whatever the differing degrees of importance attached to the broad purposes of assessment, everyone believed the work of assessment to be of high seriousness. Assessment was seen as the principle and procedures of educational verification; one could not claim that education achieved anything at all unless assessment was there to provide credible evidence of student achievement. Assessment was necessary, not merely to reassure university teachers that their activities had some effect, but to provide evidence that public money was being wisely spent. It mattered both to the academic community and to society at large that graduates were worthy of the qualifications awarded to them and that those entering the professions were fit and safe-to-practise. Almost all interviewees spoke not only of public accountability and responsible certification, but also of the seriousness of assessment within the lives of those assessed. Interviewees cared about the quantity and quality of student learning. Because assessment was a serious matter for staff, for students and for society, interviewees found the time to carry it out with a level of care and conscientiousness that satisfied them.

There is some doubt about the generalisability of this claim. Those interviewed noted that some of their colleagues regularly complained about the intolerable burden of their assessment workload; there were references to the taking of shortcuts that would be difficult to defend. In addition, those responsible for workload models and calculations were said consistently to underestimate the time that had to be devoted to assessment; less time was allowed than its seriousness merited.

Professionalism in assessment, as in anything else, will not grow and flourish unless the conditions are appropriate. The data suggested a growing concern that inadequate time available for assessment meant that the existing level of professionalism would not be maintained, let alone developed.

c. Not all assessors possessed the desirable level of assessment capability.

Proficiency in assessment requires something more than knowledge of the subject content area to be assessed. This is not as obvious as it seems. During an AWG discussion about the membership of boards of assessors, the Convener said, 'We can go on the assumption that anyone who is acceptable to the University as a teacher is automatically competent as an assessor'. This met with immediate agreement; it was a simple extension of the argument that everyone with command of the content of a subject area can be assumed to be an effective teacher within that area. Perhaps, of course, assessment needs something more, but whatever that is it can be acquired very quickly and easily. As one interviewee said, 'Academics are bright people; if there is anything to learn about teaching and assessment, they can do it very fast'.

'Assessment capability' is an unusual term and its adoption needs some justification. There were several expressions to be avoided. 'Possessing the appropriate knowledge base' suggests that traditional technical view of professionalism in which an extensive knowledge base is first acquired during a long period of study and then applied to solve some practical problem. Following Boyer's (1990) promotion of the idea of the scholarship of teaching, one might have written of 'possessing assessment scholarship'; this does seem to focus on mastery of a large volume of literature rather than on skilful deployment of understanding

derived from it. One might have written of 'reflective practice' i.e. about assessors who had some experience of assessing and who thought about what they were doing; but this seem to neglect the cognitive component (see Barnett, 1997) i.e. what people possess to reflect with. How about 'assessment competence'? But competence is derived from an analysis of the past to serve the present, and it neglects that underpinning understanding which gives the flexibility necessary to plan for and adapt to the uncertainties of the future.

Much of what interviewees had to say revealed their confidence that they had the necessary capability to assess students professionally. The exceptions were very illuminating. 'We [in chemistry] don't know enough about objective testing and computer-based assessment. People like you should put on courses for us.' 'We have precious few ideas in philosophy about new methods of assessment – perhaps the LTSN will help us.' 'I am only an acting head [of a design department]: there is a lot I don't know about assessment.' The most striking comments of all came from medicine, where there had been by far the greatest changes in assessment practice: 'Many of my colleagues don't know enough and haven't thought enough about assessment'; 'I have applied myself to the assessment literature, some colleagues haven't bothered or been able to find the time'. The single person who declined to be interviewed gave a memorable reason: 'Although I am responsible for assessment in Year 4, I have no specialised knowledge of assessment'.

A degree of confidence is justified that assessors possess the capability, the expertise, to assess students on current courses. Whether they have the knowledge and skills to initiate and to implement new assessment procedures is open to doubt. There is, at the very least, considerable scope for professional development to support those university teachers who are (or who should be) facing the challenges

of new practices in assessment. This probably does not mean the taking of lengthy courses; it does mean the provision of adequate time for staff to attend to the appropriate forms of professional development that some of them see as necessary and all would benefit from. The need for professional development in assessment was highlighted in both the Dearing and GMC Reports; it did not feature in the policy documents generated within the University.

d. Assessment practice was not the subject of much critical reflection and creative thinking.

This claim applies to three subject areas, but not to the fourth (medicine). The evidence came from three main sources. Firstly, interviewees frequently commented that the critical reflection prompted by the interview was rarely required elsewhere. Secondly, thinking and talking about assessment were confined to disagreements about individual cases and border-line decisions and did not centre on 'the big issues'. Thirdly, there had been little creative development of assessment, except when new courses had been introduced.

In three of the subject areas there was little evidence of that critical professionalism that Walker and colleagues (2001) were keen to develop within their practice. It is perhaps significant that their own focus was on purposes and the teaching methods by which these were pursued and that they had very little to say about assessment. They conclude their book with twelve questions to provoke critical reflection; none deals with assessment. It is not possible to reconstruct professionalism in university teaching without critical reflection on present assessment practices and without creative thinking about how these should be

changed. According to Barnett (1997), professionalism requires not just being competent within a paradigm, but being critical of it; in this research there was much more evidence of competence than of critical and creative thought.

e. Commitment to individual ethical action was high; there was less commitment to communication and social action.

Consider the following: those interviewed found the time in their busy lives to carry out all the business of assessment with conscientiousness and care; they worried about inconsistency in their assessments and possible assessor bias; they sought to minimise any potential injustice to students within assessment; where plagiarism was seen as likely, they wished to find ways to maintain the integrity of their assessment; they were concerned that both policy and student numbers were reducing the ways in which boards of assessors could deal with students as persons; they feared that future policy would restrict their independence and autonomy and might diminish the extent they felt trusted to balance the interests of academic standards and of students as persons. The evidence from the interviews showed groups of academic assessors whose assessment practice was embedded in a developed and demanding ethical code; as individuals they recognised the difficulties inherent in all assessment and were willing to shoulder responsibility for the potentially serious consequences of their assessment decisions.

The evidence from the data is much less convincing in relation to commitment to communication and to social action. Particularly in chemistry and philosophy, a picture emerged of assessment being characteristically a matter of the single assessor operating on his/her own on the assessment of the products of students. Assessment was something **done to** an individual by an individual, not

something done with. In these subjects, communication and social interaction seemed to be viewed either as a time-wasting distraction or a dispensable luxury; there was little sense of inter-personal interaction being a pre-requisite of effective assessment. In medicine the necessity of collaboration over some assessment methods pointed up the desirability of better communication on the other aspects. In design, there was a distinctive stress on assessment as dialogue, both between staff and students and amongst a wider community of assessors. Examples of collaboration and partnership with assessors (other than external examiners) from different institutions were rare.

If re-defined professionalism gives an important place to communication, negotiation and partnership, then commitment to communicative, social, interactive professionalism within assessment was very uneven in these four subject areas and sometimes non-existent.

SECTION 6: PERSONAL REFLECTION ON SALIENT ISSUES

In section 4 of this chapter, there was discussion of issues emerging as important to the participants and of issues predicted by the researcher to be relevant to professionalism. This seems the place to identify issues salient from a number of subject areas. These are issues which personal reflection suggests are very deserving of future analysis and debate.

1. Plagiarism

The findings showed widespread concern about plagiarism in one form or another. There was a belief that student dishonesty was increasing and had to be met with thoughtful and creative (rather than panic) measures. There was evidence,

however, of some confusion in people's thinking about what should count as dishonesty. It was clearly desirable for students to learn from printed sources (scholarship and research) and from each other (peer learning and group-work); these could not be seen as dishonest. Dishonesty entered with the intentional misrepresentation of the sources of material or ideas; it could not always be easily detected or distinguished from carelessness or incomplete understanding of the conventions of attribution.

Some of the remedies offered seemed over-simple. As examples: scope for dishonesty had been increased by advances in information technology, therefore the only effective response must lie in more use of technologically sophisticated detection procedures; non-traditional methods of assessment prompted new forms of dishonesty, therefore there should be a retreat to conventional written examinations.

Dishonesty, its nature and prevalence, deserves more careful analysis and discussion; there is need for a more energetic and coherent attempt to devise imaginative responses and to evaluate them in terms of both desirable and unintended outcomes.

2. Communication

Overall there was inadequate communication (interactive professionalism) amongst assessors. Sometimes the need for this was not recognised; sometimes it was recognised but time was unavailable; sometimes the appropriate fora did not exist; often discussion focused on individual cases and not on principles and procedures; sometimes debate in committees was not adequately communicated to non-members.

If the maintenance of an established assessment system requires effective communication, then the planning, implementation, monitoring and development of changed systems requires more (and much more effective) communication. The situation was particularly interesting in medicine. The design and planning of a new assessment system had successfully prompted a huge amount of consultation and inter-assessor communication. The effort to get the new curriculum up-and-running then seemed for a while to engender a form of communication fatigue and the need for continuing review of assessment was temporarily forgotten.

The argument that communication about assessment remains necessary should not be trivialised into the message: 'It's good to talk and chat' (although perhaps informal, spontaneous 'blethers' are necessary). There is a difference in kind between the communication amongst assessors as they walk along pavement or corridor to the coffee-room (as happened to maintain the assessment system in design) and intense, focused and deeply serious discussion on principles, procedures and issues (as was engendered in medicine). The policy-required approach to course design ensured that assessment was discussed in the context of aims and teaching/learning methods; this was distinctively different from the relatively superficial communication about assessment as an activity detached from the educational enterprise.

3. Sustainability

Not all the interviewees had 'got the message' that how students are assessed has a potent influence on what they perceive as important to learn. It may thus seem unrealistic to argue the value of grappling with the concept of sustainable

assessment, in which the key challenge is to provide assessment strategies that meet present needs and prepare students to meet their own future learning needs.

The concept of sustainability, if not the actual term, had entered the thinking of the medical course planners and some design practitioners. Although assessment at the end of the courses was focused on diagnosing fitness-to-practise, it was recognised that within-course assessment should try to encourage future learning in continuing professional development. The idea was not obvious in chemistry and philosophy, thus an interesting question arises as to what should count as sustainability in these more 'academic' subjects. Where there is no assumption of continuing-to-practise after qualification, it is less clear where future learning needs will lie. What assessment-encouraged learning is most worthwhile in helping people become 'fit-for-living'?

This writer had an 'ah-hah' moment when he first encountered sustainable assessment. If the implications of sustainability in assessment could be energetically tackled then a positive quantum-leap in assessment practice becomes a possibility. It is worth noting that a Glasgow colleague was so stimulated by meeting Boud in February 2002, that four months later he had re-designed the assessment within two degree courses. (Southern – in press.)

4. Assessment professionalism

One view of academic professionalism is that it reached a high point between 1945 and the end of the 1960s ('when everything seemed possible in a benign state and a respectful society') and declined thereafter (as it 'found a hostile public led by stridently anti-academic politicians convinced of the truth of monetarism'). The quotations are from Halsey (1992). This could lead to the

conclusion that academics are currently in a post-professional phase. There are alternative views. Perhaps academic professionalism never genuinely existed and academics are currently in a pre-professional phase. Probably it is more helpful to see academic professionalism as having declined when judged against criteria which are no longer the most appropriate ones; in this case an opportunity exists to claim a 'new' professionalism judged against different criteria.

The authors who argue that this new professionalism can be constructed around the central idea of the academic as educator (e.g. Piper, 1994, p 239 and Nixon, 1997a – p100) tend to argue throughout about teaching. The 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 are significant advantages to be gained, both for academics and their students, by insisting that student assessment be seen as a set of practices calling for professionalism. As noted several times in this thesis, routine assessment within an established status-quo requires professionalism. The devising and implementing of innovations in assessment make much greater professional demands. It is in relation to changes in assessment that academic professionalism can more often be judged as inadequate to the tasks faced. A lack of professionalism limits academic capability to generate constructive, considered and collegial assessment development in the face of those pressures that currently threaten the very idea of 'academic'. (See Taylor, 1999 – p 138 and Appendix X, pp 41-43)

CHAPTER 10

REVISITING THE THEORETICAL FRAMEWORK AND REFLECTING ON THE RESEARCH

REVISITING THE THEORETICAL FRAMEWORK

The research began with an apparently simple framework; there were three primary concepts (policy, practice and professionalism), two permeating secondary concepts (change and disciplinary difference) and six 'early-day' notions of links between the main concepts. (Pages 40-43) Preliminary exploration of the concepts made it clear that this framework was already inadequate. Having carried out the planned research, how should the framework be revised?

Concept 1 It is helpful to distinguish between policy-as-product (what the policy-making of authorised choice generates) and policy-as-process (the policy-activity intended to influence the policy generated).

Concept 2 One can conceive of 'assessment practice' as a general notion referring to the whole institution and there were elements of practice which were uniform across the whole institution. However, there were distinct differences in assessment practice; these were strongly influenced by location of the subject area on the four-quadrant framework. This framework is a relatively crude classificatory device; assessment practice in any subject area will have distinctive features. To interpret the findings of the research it seems desirable to conceive of 'local' rather than institutional assessment practice, leaving it unspecified just how limited is the locality.

Concept 3 It now seems necessary to make a clearer distinction between professionalism as something possessed and as something expressed. Assessment

practitioners can be inferred to possess some concept of academic professionalism (and what we have called assessment-professionalism); this is an organised mental schema or a general attitude to their academic work. Such an attitude has a cognitive component (what they know and understand), a values component (what is seen as worthwhile) and a tendency-to-act or behavioural component. As a second-level abstraction we can conceive of a general, located conceptualisation of professionalism, compounded of the concepts possessed by individuals.

Subject culture (of which relevant policy is but one aspect) has a powerful influence on how professionalism is expressed and made manifest; aspects of professionalism can be inhibited or encouraged by current subject-based norms and cultures. There is a sense in which professionalism as a concept possessed has a more enduring life than any particular expression of it. However, the concept itself is not static; external events, pressures and policies may well prompt re-definition i.e. re-conceptualisation.

Relationship 1 This was originally expressed as a causal statement:

‘Assessment policy is an important cause of change in assessment practice’.

Temporal precedence was implied along with constant conjunction and strength of association (i.e. more practice change with policy than with other possible causes).

The research indicated that policy-makers construed the link in this way (the university would generate policy which would bring about the desired changes in assessment practice), but over the lengthy life of the policy-making group (and the policy-activity it prompted) faith in the ‘strength of association’ aspect declined. (More change would be brought about by other causes.) From the practitioners’ perspective, policy was perceived as having been one cause amongst several (and

often a rather indirect one); it was predicted to become more significant in future. In other words, the research encouraged a move away from any simplistic view of causality to one of causal multiplexity. The findings did not deny the statement. It can stand with the emphasis on policy as **one** cause of practice change.

Relationship 2 The suggested link was ‘existing assessment practice causes the emergence of new policy’; this posited a causal relationship similar in nature to the one above. There was a perception within central management and the policy group that existing practice in some locations should be ‘brought into line’ and that policy would do this. However, new policy was prompted not only by the perceived state of existing practice within the institution but also by the requirement of national policy that a local policy be formulated. The reason for this national policy probably also lay in perceptions that an undesirable variation in the quality of assessment practice existed across the higher education sector.

There was little evidence of positive practice bringing about policy change, except within medicine where thinking about curriculum and assessment were in a very active state. The merits of new practice resulted in policy-activity to change the thrust of policy-making with regard to uniform grading scales. This can be seen as confidence in practice resulting both in policy-resistance and policy re-formulation.

In Chapter 9, section 3, the statement was elaborated to focus on **mismatch**, between existing practice and what policy-makers thought it ought to be, as the main cause of policy generation. However, the statement can stand if it emphasises practice as **one** cause of policy change.

Relationship 3 The link was given as ‘Emerging assessment policy causes a restriction of professionalism’; the nature of the claim was similar to the first two.

The balance of the evidence supported a claim that those interviewed saw a trend towards ever more policy gradually reducing the scope they had for autonomous professional judgement. There is room to doubt whether so far professionalism had in fact been restricted to the extent that was asserted. There was, for example, no obvious encroachment on academics’ responsibility for choosing how their students should be assessed or on the separate individual judgements that academics made when assessing student work or performance. Policy directives were almost entirely devoted to administrative/bureaucratic aspects of assessment. The nearest that policy got to educational aspects of assessment was in requiring that assessment procedures be explicitly aligned with intended learning outcomes.

The distinction between educational and administrative aspects is, however, suspect. Two examples must suffice. The proposal for a particular grading system was defended by policy-makers on grounds of administrative convenience, but was contested by practitioners on the grounds that it distorted perceptions of what educationally their courses were about. Secondly, the introduction of more efficient procedures into boards of examiners was said to be an administrative necessity, but it resulted in academics’ feeling that the scope for professional judgement was restricted within more mechanistic approaches.

Whether or not new assessment policy had restricted professionalism in the past there was a very strong feeling that it was going to do so in the future; the majority of people detected a worrying trend. This may have stemmed from a general concern about the worsening of conditions in higher education (more, and

more diverse, students to be given a satisfactory experience of higher education with a declining unit of resource) or from a more specific distrust of 'the central policy-makers' in relation to assessment.

There was no strong general evidence of policy having caused change in how professionalism was conceptualised. Within medicine, people were thinking about whether assessment professionalism as traditionally conceived was adequate to cope effectively with a radically new assessment system. That professionalism should be re-thought was the central argument of the book written by other academics from the same university. (Walker, 2001) That it was actually happening did not emerge clearly from this research.

It is hardly wise to make a claim to explain what has happened, when it is unclear whether it has happened or not. The explanatory claim appears plausible, but at this stage there is insufficient evidence either way. Assessment policy may affect how professionalism is manifested and may prompt its re-definition.

Relationship 4 'The greater the professionalism of academics, the greater their input to policy.' This statement made an explanatory claim of the 'biological gradient' type (the more of A, the more of B).

Suppose this were to be a formal hypothesis and the sole subject of a research project, what might the researcher do? One way of proceeding would be to estimate the level of professionalism in a number of academics and the extent of their input to policy; a positive relationship between the two measures would then provide some corroboration of their relationship.

In chemistry, only one or two interviewees had any involvement in university-level committees with a concern for assessment i.e. with any formal input

to institutional policy-making. This is not the same as saying the others were never engaged in policy activity. For example, departmental opinion had been successfully mobilised in opposition to modularisation policy. In philosophy, two people had experience of policy-making committees; there was little or no evidence of assessment policy-activity within the department. The evidence suggested that, in philosophy, policy was accepted as something which arrived from outside and had then to be implemented, perhaps with complaint, but also resigned compliance. There was some mild policy-activity within the School of Design, but little or none at institutional level. Medicine was again very different; introducing a new assessment system had sharpened awareness of the impact of existing and emerging policy on the practice they wanted to adopt. There was plenty of evidence of confident and energetic policy activity, clearly aimed at changing policy. All this is interesting, but it does not add up to strong evidence to support the ‘hunch-hypothesis’.

The writer is very much aware at this point that this is not a retrospective explanatory claim that the research supports. Rather it is a relationship that he personally would wish to be true, not ‘this was the case’ but ‘this **should** be the case’. It would be a good thing from his perspective if assessment professionalism were to become greater, and one consequence were to be a greater input by academics to policy-making and policy-activity.

Relationship 5 The grounds are firmer for statement 5: ‘The more change in assessment practice, the greater demands made on professionalism’. In philosophy, a change in assessment practice designed to reduce eccentricity in assessment had led practitioners to discuss the tension between individual freedom and group

professionalism. In chemistry and design, when there were significant changes in assessment with the introduction of new degrees, staff were challenged to think more critically about both the old and the innovative methods of assessment, to read more, to negotiate and co-operate with new assessment partners; they had become more aware of professional development needs. The same position was even clearer in medicine.

The maintenance of an assessment system does require professionalism, but increased demands are made on that professionalism whenever new practices are seriously considered. Certainty about a claim of this type is of course unachievable, but it can stand as a relationship claim with an acceptable level of validity supported by credible evidence.

Relationship 6 ‘Where there is professionalism, there is good work in assessment’, is an explanatory claim based on strength of association. It could be reworded as another biological-gradient claim: ‘The more professionalism, the better the assessment practice’.

Everyone interviewed treated assessment with high seriousness; there was strong evidence of ‘good work’ in assessment. There was professionalism in all aspects of current practice. The disturbing questions, however, are ‘Should assessment practice be enhanced beyond the “good-enough” ?’ and ‘Would greater assessment professionalism lead to better assessment?’ The evidence pointed to some lack of commitment to reflecting critically on existing practice and surprisingly little innovation in assessment, except when new degrees and curricula were introduced. If critical professionalism requires practitioners not only to be

competent within existing assessment systems, but to be critical of them and creative in changing them, then there was little critical professionalism around.

This suggests that a valid retrospective explanatory claim can be made: **‘Traditional professionalism brought about good work in assessment’**, but again the writer wants to add an additional values-claim. **‘Critical professionalism would result in better work in assessment.’**

The research data and what has been written above prompt an attempt to re-draw the theoretical framework depicted on Page 8. This is shown in Figure 4, over.

Notes on Figure 4

- The figure applies to each subject-area site within the institution studied. Theoretical inferencing implies the framework can be generalised to other similar institutions. Naturalistic generalisation is a matter for the reader.
- Although for some purposes ‘local’ assessment practice and institutional policy-making are useful reifications, the idea of a local conceptualisation of professionalism is more problematic. Individuals in the location have their own conceptions of professionalism; that there is some group consensus or one dominant conceptualisation at any given time involves an inferential leap of a size that may be indefensible.

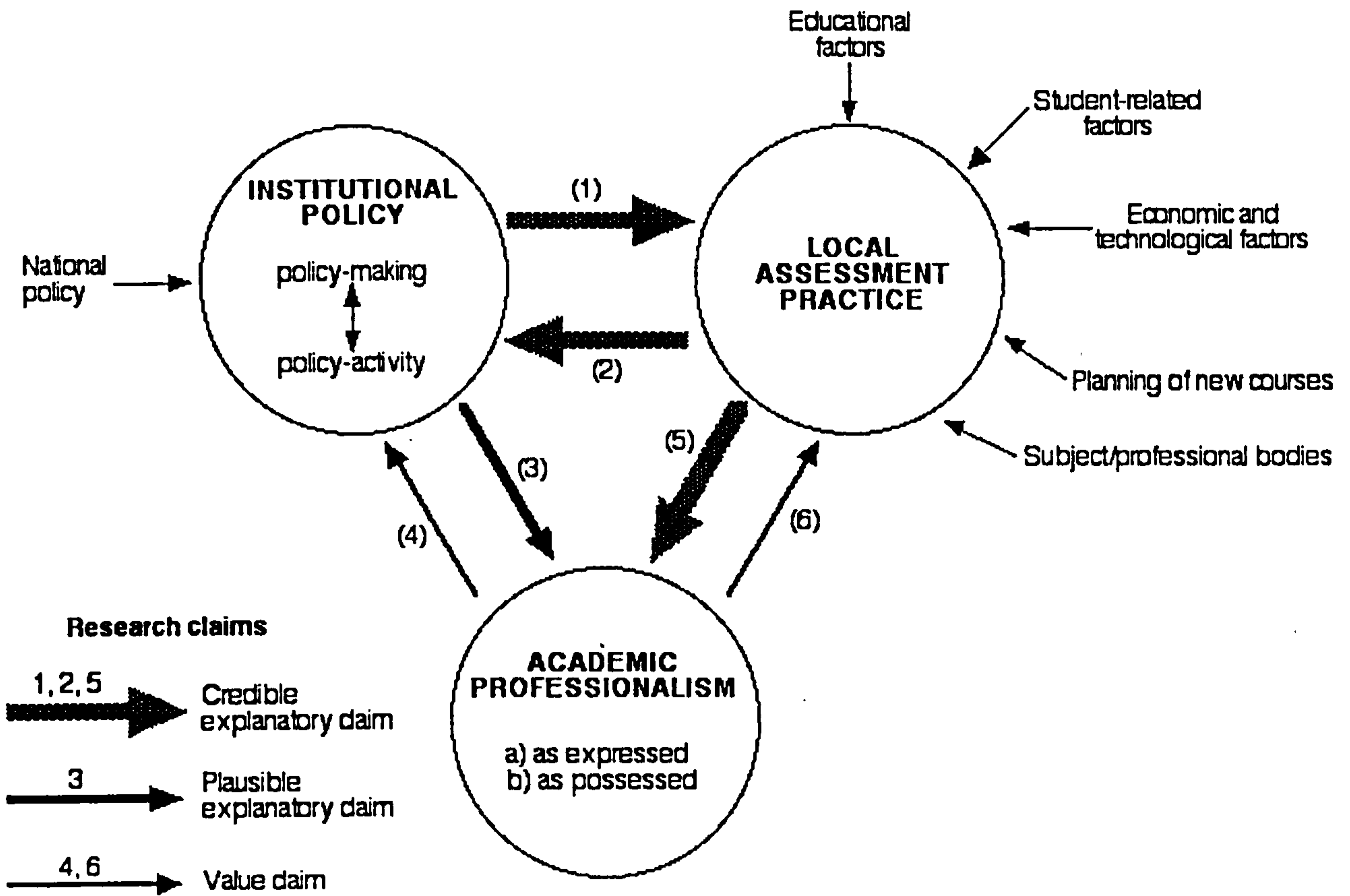


Figure 4 Relationships among institutional policy, local assessment practice and academic professionalism: a revised framework. (Compare Figure 2, Page 39)

- **Assessment practice appears on this diagram as an important nexus, or knot-concept, related to a large number of influential change factors. An alternative diagram could be drawn in which practice was centrally placed, with both professionalism and policy as change factors within a constellation of others. Change in practice has been a central interest in this research and clearly has a multi-factorial aetiology.**
- **Relationships 1, 2, 3 and 5 retain much of the character they had in the entry-diagram; they are retrospective claims about links with some explanatory potency and for which the research provided varying amounts of credible evidence.**
- **Claims 4 and 6 are ones for which the research provided insufficient evidence; research is needed which focuses selectively on these linkages. Unable to make explanatory claims which satisfied him, the author felt moved to substitute claims which had no legitimate place within the research. He could not say ‘the greater the professionalism of the interviewees, the greater was their input to policy’; he wanted to say ‘Their professionalism should have been greater and then they would have made more input to policy’. He could say ‘Professionalism resulted in good work in assessment’; he wanted to say ‘If professionalism were re-defined as critical professionalism this would result in better assessment’.**

The simple entry-level diagram served well-enough as a general guide to organise the research and to structure this report. The diagram as revised in Figure 4 retains the main features of the original; the revision has made it more complex, but

not rejected it. The hope that the original diagram represented a theoretical framework which the research could transmute into something more satisfying, elegant and significant was over-optimistic, perhaps naïve. The original framework was simplistic: the complex nature of the component concepts and the diffuse and variable character of the links between them inevitably meant that an attempt at refinement and clarification would result in complexification. This was not so much unfortunate as inevitable given the area investigated and the approach adopted. Research of this kind confirms that the causes of human actions are not single, simple and direct; they are 'multiple and conjunctural, affecting each other as well as their supposed effects'. (Huberman and Miles, 1994)

REFLECTING ON THE RESEARCH

Without assuming that the finishing post will ever be reached, it is at least in sight, so how does this researcher appraise his research? It is time to contribute personal answers, acknowledging that the judgements of others may be very different. There are three general questions. Was the research well-done? Was it worth doing? What should happen next? More specific questions will be used to structure the remainder of this chapter.

- A. Was the research well-done?
 1. Were the research purposes appropriate ones?
 2. Was the research design fit-for-purpose?
 3. How well was the research managed?
 4. How competently were the research methods conducted?

5. How appropriate, to its different audiences, were the written accounts of the research?

B. Was the research worth doing?

1. What did the researcher learn about himself?
2. How significant was the original contribution to understanding?
3. What impact did the research have on practitioners and policy-makers?
4. What might have been done differently to make the research more worthwhile?

C. What should happen next?

1. What research should this researcher do next?
2. What should this researcher write next?
3. What future research is recommended to be done by others?

A. WAS THE RESEARCH WELL-DONE?

A 1 Were the research purposes appropriate?

The researcher is still wholly convinced he was right to choose the general area of assessment in higher education. It matters (a) because of the crucial importance of assessment as an influence on student learning and (b) because assessment in higher education remains under-discussed and under-researched and is the least well-understood and enacted aspect of higher educational provision. (Wakeford, 1999 and Yorke, 2001) Growing general recognition of the importance of the area is confirmed by the amount of written material published during the research; this applies to policy products (e.g. QAA, 2000), research reports (as evidenced by the growing number of assessment entries in *Research into Higher*

Education – Abstracts - SRHE, 1998-2002), critical papers (most recently - Knight, 2002) general texts for higher education lecturers (e.g. Brown and Glasner, 1999; Heywood, 2000; Schwartz and Webb, 2002) and staff support materials (ILT on-line resources, 2000-2002, and LTSN, 2001).

There were three broad aims (Page 2). The first was about understanding why in higher education assessment things are as they are and why things change. The first element in this now appears futile in relation to interview-based research; the obvious response is ‘Things are as they are, because they are much as they always were and no-one now remembers why they came to be that way’. This is not to deny that historical research (based on the scrutiny of policy documents, committee papers, course documentation and assessment instruments) could be of value.

In the second aim, the verb ‘to explore’ (relationships among assessment policy, practice and professionalism) is weak. There seems little point in the process of exploration if there is no explicit intention to deliver some helpful product, at least a helpful map of the territory. The explorer may well enjoy him/herself, but this is not enough. A more satisfactory formulation of this purpose would have been ‘to refine simplistic assumptions about the nature of the relationships...’ The aim is still seen as appropriate for research of this type, but it may well have been too ambitious for a single, part-time researcher. The three conceptual organisers (policy, practice and professionalism) were central to the taught components of the EdD programme and perhaps this researcher was too ambitious in attempting to conduct research to say something useful about all three. The third aim (contributing to policy and practice) came straight from EdD documentation. The aim is still judged most appropriate; the researcher has provided assessment practitioners and policy-

makers with a research contribution which these people have seen as useful.

Evidence is provided in B 3.

More precise purposes were explicated in the form of research questions.

(Page 3.) These questions were appropriate in the sense that (a) they were clearly derived from the general aims and (b) they turned out in practice to be effective guides to research activity. Earlier versions were modified through discussion with key informants and pilot interviewees and also received validation from them.

Although the questions were broadly satisfactory, two in particular deserve comment.

RQ 3 This question (about participants' views of what was pleasing in their assessment practice) was very helpful in conducting the interviews. It encouraged people to talk and allayed any suspicion that the researcher might focus on deficiencies. It was decided that the answers (although a valued focus in the internal subject reports for the University) were not central to the development of the arguments in this thesis. The question has thus achieved the anomalous status of being very helpful, but only answered in Appendices.

RQ 7 The wording of this question (about the impact of policy and practice on professionalism) is open to three criticisms. Firstly, it neglects the reverse direction of 'impact', i.e. professionalism can have an impact on assessment policy and practice. Secondly, it may mislead by suggesting that the research will tackle questions about impact directly (by obtaining the participants' views on impact, or even observing it); this was not the intention and it was not done. Thirdly, if Aim 2 was too ambitious then so too was this question derived from it.

A 2 Was the research design fit-for-purpose?

Given the above purposes, the design was appropriate. The number of research questions and their heterogeneity meant a diverse range of methods were employed; the approach was necessarily eclectic. This is defensible; the methods complemented each other (allowing a modest amount of triangulation) and were consistent with the general theoretical perspective.

The researcher had a little early unease at introducing two structured episodes into semi-structured interviews; events showed this unease to be unjustified. This is not to say that the placing within the interviews was unimportant or uninteresting. This aspect provided the material for the methodological section of a presentation at an international conference. (Holroyd, 2001)

The research design, it is claimed, was fit for its purpose; it also proved feasible to implement. There were occasions when the researcher felt he had attempted too much, but the fieldwork deadlines specified in the design were actually met. A worry remains that the quantity of work resulted in some depth being sacrificed to breadth.

A 3 How well was the research managed?

It is difficult to get the tone right in a reflexive account. One has to be perceptively self-critical, without lapsing into embarrassing self-flagellation. Conversely, it is difficult to praise oneself, without sounding smug.

At the risk of immodesty, this aspect of the research is the one most deserving of praise. It still surprises the researcher that he arranged and conducted thirty-eight interviews without any administrative blunders or diplomatic incidents. He arrived in the right place, at the right time, having done his homework and

properly equipped. Those interviewed were busy people, sometimes seriously over-stretched and not always well-disposed to educational research; they were not the kind of people to tolerate mismanagement or to suffer fools gladly. The rules of engagement were carefully devised and strictly adhered to. The research was well-managed in the sense of being ethically sound.

The research design required data analysis and writing-up for one subject area at the same time as the field-work was being carried out in the next. This made heavy demands on time-management skills. If the researcher had been working full-time, his management skills would have been stretched to breaking-point.

The Vice-Principal (Learning and Teaching) and the Director of the Teaching and Learning Service, neither of whom is normally effusive, have praised the management of the research and indicated that its conduct has made it more likely that such research will be permitted, indeed encouraged, in future. It is possible that their main reason is relief that the researcher caused them no trouble. This leaves the researcher with the feeling that he may have been unnecessarily emollient. If occasionally he had been more intellectually assertive and confrontational he might have been judged less good as a manager.

A 4 How competently were the research methods conducted?

The form of the question assumes that there are levels of competence; this will not be examined here, but see Holroyd (1999 b) and Elliott (1991). Rather than pronounce himself 'competent', the writer attempted as much detachment as possible by evaluating himself with the Dreyfus (1981) five-stage model of developing competence. The five stages are novice, advanced beginner, competent, proficient and expert; key criteria centre on the nature of component, salience and

whole-situation recognition and the decision-making displayed. There is another academic paper here, but summary judgements only will be given.

1. Participant observation Evaluation: progress from 'advanced beginner' to 'competent'.

In his involvement with the policy-generating group, the researcher had difficulties resolving the role-tension between engaged member and detached observer over the very long life of the group. Because of uncertainty as to how the study of the group would contribute to the purposes of the research, there were problems with salience recognition. ('What am I here for?' 'What is it most important to record?') When the purpose became clear, the conduct of the method became competent.

2. Interviewing Evaluation: generally proficient; a few lapses into merely competent; expert on occasion.

Reflecting on his interviewing with the help of transcripts and tapes, the researcher was generally satisfied. There were a few times when he regretted his failure to pursue and probe interesting points. There were also occasions when he was pleasantly surprised: 'I don't know why I did that, but it worked.' If this is what Dreyfus means by decision-making going beyond the rational to the intuitive, then it was the performance of an expert interviewer.

3. Documentary and data analysis Evaluation: competent.

The researcher wishes to believe the various analyses were competently done; he can claim no more from a Dreyfus approach. However, and probably as a form of ego-defence, he is unconvinced that Dreyfus is helpful in this regard. In these forms

of analysis there is an obvious place for holistic whole-situation recognition and for intuitive decisions (What is the likely sub-text?), but he cannot conceive of forms of analysis which can not be described as analytical. There is something wrong with the Dreyfus model if no-one can be judged an expert analyst.

A 5 How appropriate, to its different audiences, were the written accounts of the research?

There were four distinct audiences.

1. The research participants and their subject-area colleagues.

Draft reports were sent to all interviewees. There were very few suggestions as to how the reports should be revised to communicate more effectively. The four revised subject reports were well-received; all the relevant heads of department decided to circulate them beyond participants.

I attach my comments on your absolutely fascinating document. Reading it left me with a real feeling of admiration both for your tenacity and for the assessment processes of my studio colleagues. This document has spurred me on to reflect in various ways on what we do in this department and to encourage the valuing and development of alternative critical submissions and a more evolutionary view of student performance. Thank you for this opportunity to think very hard about a central aspect of our work. I assume the report is not confidential, as I would very much like other members of the staff to read and discuss it.

[Head of section]

2. The Vice-Principal (Learning and Teaching) and the Director of the Teaching and Learning Service.

The subject reports were submitted to the Vice-Principal. He found them 'interesting and useful, but too long for my purposes' and requested a summary report. This was written, judged appropriate and passed to the University Education Committee and subsequently to the Learning and Teaching Committees. The

Director of TLS pronounced the five reports helpful to the work of the service; parts of them would be used in future educational development work.

3. The wider academic community.

A paper was submitted to the journal *Active Learning in Higher Education*. The two referees recommended it for immediate publication. Only one change to the submitted text was suggested; this meant providing two additional references for a claim made about the assessment literature.

4. Examiners appointed by the University of Stirling.

The writer has tried to convince many students that being succinct is a virtue; he has found it extremely hard to follow his own advice. ('I can easier teach twenty what were good to be done, then be one of the twenty to follow mine own teaching.') Perhaps because he did too much research, he found it profoundly irksome to conform even to the extended word-limit granted him. He is tempted to argue that there should be no word-limit for doctoral theses, that candidates should only be penalised if the text is over-aerated. The quotation above from the Merchant of Venice is an intentional example of such over-aeration.

B. WAS THE RESEARCH WORTH DOING?

B 1: What did the researcher learn about himself?

It is difficult to disentangle new learning from confirmation of what was already suspected; what follows is a mixture.

(a) Memory and concentration span are declining with age.

(b) The researcher needs sympathetic but critical colleagues with whom to talk.

The production of language does not merely reveal thinking, it develops it.

(c) It was easier for the writer to see himself as a junior research worker than it was initially for some of his interviewees. Uncertainty as to just who the researcher was may have affected the introductory phase of some interviews. The transcripts and tapes have been revisited and the researcher would claim that, once the interest of the task was established, interviewee perceptions of the interviewer's role had little influence on what was said and posed minimal threat to the validity of claims made from the data. (See also Pages 60-61.)

(d) He is still a good listener. He had not previously conducted research interviews with staff in higher education and was unsure if he had the skills necessary to deal with hard-pressed, highly-intelligent 'subjects'. All interviewees required him to be wide-awake, but no-one was 'difficult'. Most of the interviewees did seem over-worked; they also wanted, perhaps needed, to talk at length. The researcher was surprised how often he was tempted to move from research interviewing into the counselling mode of therapeutic interviewing. He briefly succumbed to temptation in six interviews.

(e) His patience with committee work has diminished. Those who set up committees should think harder about their remit and membership; those who chair committees should submit to regular re-training; committee members need guidance on how to contribute economically, to maximum positive effect.

(f) As a result of reading widely and wildly for the research, he has changed his outlook on educational texts. In the past, he too often confused the well-presented with the worthwhile; he is now more confident in judging quality. Although he has become more dismissive of the third-rate, he is even more in awe of the best.

(g) Planning, managing and conducting this research were very pleasing; writing-up this formal account, on top of the earlier reports for Glasgow University, was not. Never again will he write a long document in the third person. He is also reminded of how easy it is for him (and too many others in education) to write in persuasive/ hortatory mode and how hard is the discipline of fidelity to data.

B 2: How significant was the original contribution to understanding?

(a) Was there anything original?

The research asked questions which were original. An early literature search using 'assessment /higher education' revealed thousands of references; a search on the basis of 'professionalism/ higher education' gave hundreds of references; a search with key words 'assessment/professionalism/higher education' yielded nothing. (Interestingly a book exists that such a search **ought** to reveal: 'Are professors professional?' (Piper, 1994) The title is misleading; the central concern is the professionalism of external examiners.) In addition, when an article with the title *Are assessors professional?* was submitted for publication, both peer-reviewers said this was a paper that asked new questions. The questions, and thus the answers, were **original**; this does not mean the questions were worth asking or the answers credible.

The Koestler (1964) position on originality is still helpful. All acts of creation (i.e. that produce something original), come from the bisociation of previously unrelated matrices, where a matrix is taken to be an organisation of information. In this research there were three matrices of this kind and questions were asked that required pairs of these to be brought into bisociation. This looks promising, but it has then to be asked what the nature of the bisociation was. For

Koestler, theoretical advance comes from the effective integration of matrices, but the explanation of humour lies in their collision. Thus this research may have aimed for original theory and produced only a rather bad, long-winded joke.

(b) Was there anything original that aided **understanding**?

If this research were mere data-gathering, it might have generated original information which contributed nothing to anyone's understanding.

Whose understanding? The researcher's own understanding of assessment increased, but so did that of others. Several interviewees said that their understanding had increased through their participation and the published paper resulted in positive comments from referees and some readers about its contribution to their understanding. But what is to count as understanding? In Appendix III, it was argued that understanding is more than knowledge because of its relationship with explanation. Understanding is shown when the possessor can provide satisfying and satisfactory explanations and any such explanation depends on being able to make appropriate use of theory statements which link relevant concepts. Thus we should ask whether there is any original theory in this research. Figure 4 is more complex than Figure 2 (i.e. it does have original bits in it). This complexity was prompted and confirmed by the research, but is somewhat unsurprising. It provided the researcher with little of that 'peculiar zing' of pleasing original theory.

(c) What original contribution to understanding is claimed?

Using the relatively rigorous definition of understanding developed above, the researcher claims to have made a contribution of modest significance to our understanding of how policy, practice and professionalism are related with regard to

assessment. He hoped to achieve more, but expectations may have been unreasonably high.

With a less 'theoretical' meaning attached to understanding, a less modest claim can be made. This research has added to our understanding of how academics can be encouraged to engage with important ideas. The research explored the process of getting people to give attention to assessment and professionalism and the 'real' or imagined links between them. The research achieved something that was not achieved by setting up a policy committee. Policy alone did not produce positive and productive engagement of minds. There are some limited ways in which enhancement of assessment practice can be brought about directly using the pressures and directives of policy, but more fundamental change can be stimulated by research into practice. An invitation from the researcher to engage with him in thinking about current practice led to more informed criticism of the present and to more motivation to change it. Whereas being told what to do encourages resistance and policy-avoidance strategies, being invited to think about what one does can stimulate advance.

This contributes to understanding of what can be meant by that currently popular term 'research-informed practice'. Research can produce findings which may be applied within practice; researching into one's own practice may lead to change in that practice; becoming engaged in thinking about practice by participation in someone else's research may enhance that practice more than responding to the pressures of policy.

It is right that this thesis be judged in terms of its theoretical adequacy. (Do the claims within the interpretive superstructure have any explanatory validity?) Perhaps it should also be judged in terms of its **catalytic validity** (Lather, 1991)

Have those involved in the research (the researcher, the participants and the readers) gained in understanding of their own practice and of themselves? Such a criterion seems congruent with the purposes of a professional doctorate. The evidence of the next section suggests that this criterion may have been met.

B 3 What impact did the research have on practitioners and policy-makers?

The research had much greater positive impact, in the short-term, on both practitioners and policy-makers than the researcher ever dared hope. (The long-term impact is unknowable.) The evidence for this was unsought, clear and robust.

(a) The research raised assessment higher on the agenda of many practitioners.

Assessment is one of those things that get more interesting as you think and talk about them.

(b) The research prompted some practitioners to identify professional development needs.

I have just realised that I thought multiple-choice questions were the only form of objective testing. Why don't you lot run a course on this for us?

(c) The research stimulated departments to give assessment policy and practice a higher priority in staff meetings.

Reading your draft report has made me realise the number of important underlying issues that we have probably been remiss in not addressing within this department. It is true that most of the time we spend in discussing assessment is to do with tricky decisions about individual students. This has not seemed to leave much time over for addressing principles and desirable changes. I will be making arrangements to ensure that assessment is regularly discussed at departmental meetings. [Head of department]

(d) In one subject area a new senior assessment post was created and an assessment strategy review group established.

Shortly after, and prompted by, the production of the report for medicine, the Associate Dean for Education appointed a new professor to be his assistant with special responsibility for assessment. One part of her remit was the setting up of an Assessment Review Group; the researcher was then invited to join that group as ‘an objective assessment professional’.

(e) The research had some impact in the University on practitioners outwith the original subject areas.

The researcher was invited to hold three research seminars on his work; these were well-attended and favourably received.

(f) The summary report was fed into the policy-making committee structure of the University and that structure was itself revised.

The short report requested by the Vice-Principal was considered by the Education Committee of Senate. This committee decided that there should be new faculty-based learning and teaching committees, that their remit should explicitly cover assessment and that one or more of the research reports should be discussed within them. The research was not the only cause of the change in committee structure and remits, but it is credited with acting as a catalyst.

(g) Parts of the various reports are to be fed into educational development courses for assessment practitioners.

(h) The research has had some impact beyond the University.

The journal article based on reading for the research (Holroyd, 2000) resulted in six sets of correspondence; the conference presentation (Holroyd, 2001) resulted in five e-mail interactions with academics as far afield as Australia and the South Pacific.

B4 What might have been done differently to make the research more worthwhile?

Nothing went seriously wrong in this research and, given its aims, the writer has no strong conviction that he should have done anything very differently. The experience of this research has, however, prompted him to reflect upon his research priorities. Put simply, is he a person for whom developing a theoretical framework is paramount or one who believes that, however indirectly, research must contribute to 'making things better', in this case for the primary clients of academic professionals i.e. the students?

At times the desire for theoretical respectability was uppermost in the researcher's mind and this may have led to an over-emphasis on the role of theory as explanation. It seems desirable to consider what might have happened if he had allowed passionate concern more influence on the design of the research. His concern was that there are some students in higher education who are not as well-served by their assessors as they deserve. If that concern had been allowed to remain dominant, a different type of research would have resulted. It would have focused on student experience, including bad experience, it would have been less comfortable and perhaps more challenging. It would have been more difficult to get approval for such research and also for future research.

After exploring the metaphors of conceptual basis and superstructure, it was suggested that one might need a third metaphor, that of conceptual shrubbery.

Research students may have to do something dry and academic and of little practical consequence. They may have to learn to hide their real concerns in the shrubbery of current concepts. (Fletcher, 2002)

C WHAT SHOULD HAPPEN NEXT?

C 1 What should this researcher do next?

In Appendix I, it was proposed that this interview-based study should be developed into a one-site case study of the University as a whole. The researcher would stay with the modified theoretical framework, seek more evidence for the component conceptual relationships and further refine the concepts and their links. The research questions would stay essentially the same, with minor modifications.

This remains a possibility. However, given the priorities mentioned above, he is attracted to the idea of providing guidance to the University on how to get more people thinking about the enhancement of assessment practice through research of a variety of different types. He now has a practical ambition to promote assessment thinking across the institution.

C 2 What should this researcher write next?

There is much material from this study ripe for development into academic and theoretical papers for publication and conference presentation. There is a need for materials to promote assessment-professionalism, particularly through research-participation. There is also a need for an accessible guide for researchers on the role of theory in research and he would like to write it. (Conclusion of Appendix III)

C 3 What future research is recommended to be done by others?

The writer believes that this research opens up innumerable avenues for further research. Here are eight quick suggestions:

- critical policy research into the generation of assessment policy in Glasgow University and reactions to it;
- extension of this interview-based study to other subject areas within Glasgow University, possibly using innovative forms of interviewing;
- extension of this type of research to other institutions of higher education in Scotland and beyond in the interests of generalisability;
- tackling the same practitioner-professionalism focus using a different research approach, perhaps employing observational methods;
- action research focusing on the development of assessment practice;
- research concentrating on the student perspective on assessment;
- studies probing imaginatively into what goes wrong in assessment;
- research using the concepts of single great thinkers, perhaps into the role of power within assessment processes.

REFERENCES: *MAIN LIST

- ASSHE (1996) *The ASSHE Inventory: Changing Assessment Practices in Scottish Higher Education*, Edinburgh: The University of Edinburgh and Napier University.
- Atkins, M.J., Beattie, J. and Dockrell, W.B. (1993) *Assessment Issues in Higher Education*, University of Newcastle for Employment Department Group.
- Ball, S. (1994) *Education Reform: A Critical and Post-structural Approach*, Buckingham: Open University Press.
- Barnett, R. (1997) *Higher Education: A Critical Business*, Buckingham: The Society for Research into Higher Education and Open University Press.
- Bassey, M. (1999) *Case Study Research in Educational Settings*, Buckingham: Open University Press.
- Bauman, Z. (1999) *In Search of Politics*, London: Polity Press.
- Becher, T. (1989) *Academic Tribes and Territories: Intellectual enquiry and the culture of disciplines*, Buckingham: The Society for Research into Higher Education and Open University Press.
- Becher, T. and Trowler, P.R. (2001) *Academic Tribes and Territories: Intellectual enquiry and the culture of disciplines* (Second edition) Buckingham: The Society for Research into Higher Education and 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.
- Biglan, A. (1973) 'The characteristics of subject matter in different subject areas', *Journal of Applied Psychology*, 57(3): 195-203.
- BOLOGNA (1999) Joint Declaration of the European Ministers of Education, www.med-net.nl/topics/news/bologna.htm
- Borland, K.W. (2002) 'Towards a Culture of Assessment' in Schwartz, P. and Webb, G (eds), *Assessment: Case Studies, Experience and Practice from Higher Education*, London: Kogan Page.
- Bottery, M. (1998) *Professionals and Policy Management Strategy in a Competitive World*, London: Cassell.

* This list gives references only to sources directly referred to in the main text. There are additional lists of references in some appendices (especially Appendices III and X). These latter sources may have informed the discussion in the main text

- Boud, D., Keogh, R. and Walker, D. (1985) *Reflection: Turning Experience into Learning*, London: Kogan Page.
- Boud, D. and Falchikov, N.J. (1989) 'Quantitative studies of student self-assessment in higher education: a critical analysis of findings', *Higher Education*, 18, 529-549.
- Boud, D. (2000) 'Sustainable assessment: rethinking assessment for the learning society', *Studies in Continuing Education*, 22(2): 151-167.
- Boud, D. (2002) *Towards Sustainable Assessment*, Lecture/Seminar at University of Glasgow, February 2002.
- Boyer, E.L. (1990) *Scholarship Reconsidered: Priorities of the Professoriate*, San Francisco: Carnegie Foundation for the Advancement of Teaching/Jossey-Bass.
- Bridges, P., Bourdillon, B., Collymore, D., Cooper, A., Fox, W., 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.
- Broadfoot, P.M. (1996) *Education, Assessment and Society*, Buckingham: Open University Press.
- Brown, G., Bull, J. and Pendlebury, M. (1997) *Assessing Student Learning in Higher Education*, London: Routledge.
- Brown, S. and Glasner, A. (eds) (1999) *Assessment Matters in Higher Education*, Buckingham: The Society for Research into Higher Education and Open University Press.
- Clark, B. (1987) *The Academic Profession*, Berkeley CA: University of California Press.
- Cohen, L., Manion, L. and Morrison, J. (2000) *Research Methods in Education* (5th Edition), London: Routledge/Falmer.
- Colebatch, H.K. (1998) *Policy*, Buckingham: Open University Press.
- Crotty, M. (1998) *The Foundations of Social Research: Meaning and Perspective in the Research Process*, London: Sage.
- Drever, E. (1995) *Using Semi-Structured Interviews in Small-Scale Research*, Edinburgh: Scottish Council for Research in Education.
- Dreyfus, S.E. (1981) *Four Models v Human Situational Understanding: Inherent Limitations on the Modelling of Expertise*, US Air Force Office of Scientific Research. (Discussed in Elliott, q.v.)

- Dunne, J. (1993) *Back to the Rough Ground: Practical Judgement and the Lure of Technique*, Indiana: University of Notre Dame Press.
- Elliott, J. (1991) *Action Research for Educational Change*, Milton Keynes: Open University Press.
- Fletcher, C. (2002) 'Supervising activists for research degrees: responsibilities, rights and freedoms', *Active Learning in Higher Education*, 3(1): 88-103.
- Fontana, A. and Frey, J.H. (1994) 'Interviewing: The Art of Science', in Denzin, N.K and Lincoln, Y.S. (eds), *Handbook of Qualitative Research*, London: Sage.
- 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.
- Fullan, M.G. (1991) *The New Meaning of Educational Change*, London: Cassell.
- Gibbs, G. (1995) *Assessing Student Centred Courses*, Oxford: Oxford Centre for Staff Development.
- Gipps, C.V. (1994) *Beyond Testing: Towards a theory of educational assessment*, London: Falmer.
- GMC (1993) *Tomorrow's Doctors*, London: General Medical Council.
- Halsey, A.H. (1992) *The Decline of Donnish Dominion*, Oxford: Oxford University Press.
- Hammersley, M. (1996) *Educational Research in Action (The E 835 Study Guide)* Milton Keynes: The Open University.
- Harlen, W. (1994) *Enhancing Quality in Assessment*, London: Paul Chapman.
- Harlen, W. and Holroyd, C. (1995) *Confidence and Understanding in Teaching Science in Primary School*, Edinburgh: Scottish Council for Research in Education.
- Heywood, J. (2000) *Assessment in Higher Education: Student Learning, Teaching Programmes and Institutions*, London: Jessica Kingsley.
- Hirst, P.H. (1974) *Knowledge and the Curriculum*, London: Routledge.
- Holroyd, C. (1999 a) 'Academics responding to change: an extended review', *Teaching in Higher Education*, 4(3): 429-432.

- Holroyd, C. (1999 b) 'Teacher Competence', in Bryce, T.G.K. and Humes, W.M. (eds) *Scottish Education*, Edinburgh: Edinburgh University Press.
- Holroyd, C. (2000) 'Are assessors professional? Student assessment and the professionalism of academics', *Active Learning in Higher Education*, 1(1): 62-78.
- Holroyd, C. (2001) 'Changes in student assessment: exploring the rationale', Paper presented to International Conference, HECU 2, Lancaster: University of Lancaster.
- Hounsell, D. and McCulloch, M. (1999) 'Assessing Skills in Scottish Higher Education', in Dunne (ed) *The Learning Society: International Perspectives on Core Skills in Higher Education*, London: Kogan Page.
- Huberman, A.M. and Miles, M.B. (1994) 'Data Management and Analysis Methods', in Denzin, N.K. and Lincoln, Y.S. (eds) *Handbook of Qualitative Research*, London: Sage.
- ILT (2000) see www.ilt.ac.uk/about.html
- ILT (2001, 2002) see www.ilt.ac.uk/redirectPortal.asp?article
- Jackson, N. and Lund, H. (eds) (2,000) *Benchmarking for Higher Education*, Buckingham: The Society for Research into Higher Education and Open University Press.
- Knight, P.T. (2002) 'Summative Assessment in Higher Education: practices in disarray', *Studies in Higher Education*, 27(3): 275-286.
- Kvale, S. (1996) *Interviews*, London: Sage.
- Koestler, A. (1964) *The Act of Creation*, London: Hutchinson.
- Kolb, D.A. (1981) 'Learning styles and disciplinary differences', in Chickering, A. (ed) *The Modern American College*, San Francisco CA: Jossey-Bass.
- Kuhn, T.S. (1977) *The Essential Tension*, Chicago IL: University of Chicago Press.
- Lather, P. (1991) *Getting Smart: Feminist Research and Pedagogy with the Post-Modern*, London: Routledge.
- LTSN (2001) *Assessment Series*, York: The Learning and Teaching Support Network Generic Centre. See also www.ltsn.ac.uk
- Mason, C. (1998, 2000) 'Subject differences in degree assessment', *A series of unpublished papers*, St Andrews: University of St Andrews.

- Miller, A.H., Imrie, B.W. and Cox, K. (1998) *Student Assessment in Higher Education: a handbook for assessing performance*, London: Kogan Page.
- Miller, H. (1995) 'States, Economies and the Changing Labour Process of Academics', in Smyth, J. (ed) *Academic Work*, Buckingham: The Society for Research into Higher Education and Open University Press.
- Munn, P. and Holroyd, C. (1989) *Pilot School Boards: Experiences and Achievements*, Edinburgh: Scottish Centre for Research in Education.
- NCIHE (1997) *Higher education in the learning society: National Committee of Inquiry into Higher Education (The Dearing Report)*, London: NCIHE.
- Nisbet, J. (1993) *Assessment and Curriculum Reform*, Paris: Organisation for Economic Co-operation and Development.
- Nixon, J. (1997 a) '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., Martin, J., McKewon, P. and Ranson, S. (1997 b) '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.
- Nixon, J., Beattie, M., Challis, M. and Walker, M. (1998) 'What does it mean to be an academic? A Colloquium', *Teaching in Higher Education*, 3(3): 277-298.
- OU (1996) *Educational Research in Action (The E 835 Study Guide)*, Milton Keynes: The Open University.
- Ozga, J. (2000) *Policy Research in Educational Settings: Contested Terrain*, Buckingham: Open University Press.
- Pantin, C.F.A. (1968) *The Relations between the Sciences*, Cambridge: Cambridge University Press.
- Piper, D.W. (1994) *Are Professors Professional? The Organisation of University Examinations*, London: Jessica Kingsley.
- Powney, J. and Watts, M. (1987) *Interviewing in Educational Research*, London: Routledge and Kegan Paul.

- QAA (1999) *Code of Practice for the Assurance of Academic Quality and Standards in Higher Education: External Examining – Guidelines*, London: The Quality Assurance Agency for Higher Education.
- QAA (2000) *Code of Practice for the Assessment of Students*, London: The Quality Assurance Agency for Higher Education.
- Rowntree, D. (1977) *Assessing Students – How Shall We Know Them?* London: Harper and Row.
- Said, E. (1993) *Representations of the Intellectual: the 1993 Reith Lectures*. (as quoted in Walker (ed) (2001) *Reconstructing Professionalism in University Teaching*, Buckingham: The Society for Research into Higher Education and Open University Press.)
- Samuelowicz, K. and Bain, J.D. (2002) 'Identifying academics' orientation to assessment practice', *Higher Education*, 43: 173-201.
- Schmidt, H.G., Norman, G.R. and Boshuizen, H.P.A. (1990) 'A cognitive perspective on medical expertise: theory and implications', *Academic Medicine*, 65(10): 611-621.
- Schön, D. (1983) *The Reflective Practitioner*, New York: Basic Books.
- Schön, D. (1987) *Educating the Reflective Practitioner*, London: Jossey-Bass.
- Schwartz, P. and Webb, G. (2002) (eds) *Assessment: Case Studies, Experience and Practice from Higher Education*, London: Kogan Page.
- SHEFC (2002) *An enhancement-led approach to quality assurance, Consultation HEC/02/2002*, Edinburgh: Scottish Higher Education Funding Council.
- Smith, H., Armstrong, M. and Brown, S. (eds) (1999) *Benchmarking and Threshold Standards in Higher Education*, London: Kogan Page.
- Southern, G. (In press) 'Encouraging students to develop learning skills through formative assessment.'
- SRHE (1998 – 2002) *Research into Higher Education: Abstracts*, London: The Society for Research into Higher Education.
- Stake, R.E. (1995) *The Art of Case Study Research*, London: Sage.
- Stefani, L.A.J (1998) 'Assessing in Partnership with Learners', *Assessment and Evaluation in Higher Education*, 23(4): 339-350.
- Taylor, P.G. (1999) *Making Sense of Academic Life: Academics, Universities and Change*, Buckingham: Open University.

- Toohey, S. (1999) *Designing Courses for Higher Education*, Buckingham: The Society for Research into Higher Education and Open University Press.
- Trowler, P. (1998) *Academics Responding to Change: New Higher Education Frameworks and Academic Cultures*, Buckingham: The Society for Research into Higher Education and Open University Press.
- Verma, G.K. and Mallick, K. (1999) *Researching Education: Perspectives and Techniques*, London: Falmer.
- van der Vleuten, C. (2000) 'Clerkship assessment assessed', *Medical Teacher*, 22: 592-600.
- Wakeford, R. (1999) 'Principles of Assessment' in, Fry, H., Ketteridge, S. and Marshall, S. (eds), *A Handbook for Teaching and Learning in Higher Education* London: Kogan Page.
- Walker, M. (ed) (2001) *Reconstructing Professionalism in University Teaching: Teachers and Learners in Action*, Buckingham: The Society for Research into Higher Education and Open University Press.
- Wolf, A. (1995) *Competence-Based Assessment*, Buckingham: Open University Press.
- Yanow, D. (1996) *How does a policy mean?* Washington: Georgetown University Press.
- Yorke, M. (1998) 'The Management of Assessment in Higher Education'. *Assessment and Evaluation in Higher Education*, 23(2): 101-116.
- Yorke, M. (2001) *Assessment: A Guide for Senior Managers, Assessment Series No 1*. York: Learning and Teaching Support Network Generic Centre.
- Yorke, M., Bridges, P. and Woolf, H. (2000) 'Mark distributions and marking practices in UK higher education', *Active Learning in Higher Education*, 1(1): 7-27.