

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/263088354>

Developing assessment in higher education: a practical guide

Book · January 2007

CITATIONS

64

READS

875

2 authors:



Pete Boyd

University of Cumbria

15 PUBLICATIONS 197 CITATIONS

SEE PROFILE



Sue Bloxham

University of Cumbria

24 PUBLICATIONS 434 CITATIONS

SEE PROFILE

Developing Effective Assessment in Higher Education: a practical guide

*Sue Bloxham
Pete Boyd*



Open University Press

Open University Press
McGraw-Hill Education
McGraw-Hill House
Shoppenhangers Road
Maidenhead
Berkshire
England
SL6 2QL

email: enquiries@openup.co.uk
world wide web: www.openup.co.uk

and Two Penn Plaza, New York, NY 10121-2289, USA

First published 2007

Copyright © Sue Bloxham and Pete Boyd 2007

All rights reserved. Except for the quotation of short passages for the purposes of criticism and review, no part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form, or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher or a licence from the Copyright Licensing Agency Limited. Details of such licences (for reprographic reproduction) may be obtained from the Copyright Licensing Agency Ltd of Saffron House, 6-10 Kirby Street, London, EC1N 8TS.

A catalogue record of this book is available from the British Library

ISBN 13: 9780 335 221073 (pb) 9780 335 221080 (hb)
ISBN 10: 0335 221 076 (pb) 0335 221 084 (hb)

Library of Congress Cataloguing-in-Publication Data
CIP data applied for

Typeset by YHT Ltd, London
Printed by Printed in Poland by OZ Graf. S. A.
www.polskabook.pl

The McGraw-Hill Companies

Contents

<i>Contributors</i>	ix
<i>Acknowledgements</i>	x

PART 1: INTRODUCTION AND CONTEXT	1
---	----------

1 Introduction	3
The primacy of assessment	3
Assessment pressures and influences	3
Policy	4
Research evidence	5
Why another assessment book?	6
Situated practice	7
Audience	8
Plan of the book	8
How to use this book	10
A note about terminology	13
Conclusion	14
2 The evidence base for assessment practice in higher education	15
The relationship between assessment and learning	16
The relationship between feedback and learning	20
The validity of assessment tasks	24
Constructive alignment	27
Conclusion	29
3 The conflicting purposes of assessment	31
Validity	34
Reliability	38
Effectiveness	40
Comparability and consistency	40
Equity	41
Practicability	42
Transparency	42
Attribution	44
Conclusion and possible ways forward	44

PART 2: DEALING WITH ASSESSMENT	47
4 Teaching a module: maximising the potential of assessment	49
Assessment basics	49
Assessment and learning	52
Writing assessment tasks	54
Writing examination papers	54
Designing and setting up group assignments	56
Writing assessment criteria and standards	59
Peer and self-assessment	62
Designing tasks to reduce plagiarism	63
When problems crop up	64
Conclusion	65
5 Preparing students for assessment	66
Introduction	66
What is involved in becoming academically literate?	66
A framework for preparing students for assessment: integrating guidance and feedback	71
Giving explicit information to students	72
Telling what cannot be told: tacit information transfer	74
Conclusion	79
6 Marking	81
Introduction	81
Marking and marking practices	82
Assessment criteria, grade descriptors and marking schemes	87
Other approaches to improving accuracy of marking	93
Managing your marking	95
Issues in marking	95
Peer marking	97
Marking group assignments	97
Conclusion	102
7 Providing effective feedback	103
Introduction	103
Timely feedback	104
Written feedback	104
Formative feedback	111
Engaging students with feedback	114
Evaluating the student perspective	116
Conclusion	116

8 Moderation: towards greater consistency in marking	117
<i>Marion Nuttall</i>	
Setting the context	117
Moderation	118
Internal moderation	120
External moderation	123
Conclusion	128
9 Managing assessment in different roles	129
Introduction	129
Promoting equity	130
Tutor level	130
Module leader level	131
Programme leader level	132
Head of department level	133
Chairs of assessment boards	134
Institutional leadership	135
A note on data protection	137
Conclusion	138
10 Supporting widening participation through assessment	139
Introduction	139
Widening participation, retention and achievement	140
Using assessment to support retention and achievement	141
Equal opportunities in assessment practice	143
Assessing disabled students and those with specific learning difficulties	146
Assessment and promoting race equality	151
Conclusion	154
PART 3: DEVELOPING ASSESSMENT	155
11 Planning a programme assessment strategy	157
The importance of a programme approach	157
Validation	158
Six steps to an effective programme approach	159
Conclusion: an iterative approach	175
12 Planning the assessment for a module	176
Introduction	176
Reference points	178
Deciding on the assessment plan	181
Module descriptor template	187

viii DEVELOPING EFFECTIVE ASSESSMENT IN HIGHER EDUCATION

Building in flexibility	188
Conclusion	188
13 Diversifying assessment methods	189
Introduction	189
Why diversify?	189
Caution	191
Assessment methods	203
Sources of information	204
Conclusion	206
14 Computer-aided assessment	207
<i>Sarah Chesney and Andrew Ginty</i>	
Introduction	207
Benefits of computer aided assessment	209
Challenges of computer aided assessment	210
Online testing	211
Assessing online discussion	213
Group peer assessment software	215
Supporting diverse learners	218
Cheating and plagiarism detection	218
Sources of further information	219
Conclusion	220
15 Professional development and assessment	221
Introduction	221
Workplace learning	221
Areas for review	226
Conclusion	230
<i>Glossary</i>	232
<i>References</i>	237
<i>Index</i>	257

Contributors

Sue Bloxham is Professor of Academic Practice and Head of the Centre for Development of Learning and Teaching at the University of Cumbria, England. She has been experimenting with and researching improvements to assessment in higher education for twenty years both as a lecturer, and more recently, as an educational developer. Her research interests are also in the area of supporting student transition.

Peter Boyd is Senior Lecturer in the Centre for Development of Learning and Teaching at the University of Cumbria. His background is in Education and his current teaching role is mainly within a postgraduate course for newly appointed lecturers. Peter's research interests are in the area of induction and workplace learning of newly appointed academics in professional subject disciplines and also in peer review of teaching among academics.

Sarah Chesney is Senior Lecturer in E-learning in the Centre for Development of Learning and Teaching at the University of Cumbria. Sarah's background is in using the internet to enhance education. Her MEd from University of Sheffield in e-learning cultivated her interest in networked communication. Her research and development interests are in e-portfolios for CPD, and the enhancement of teaching and learning through web tools and new technologies.

Andrew Ginty is currently Director of the Human Anatomy Learning and Research Centre within the School of Dentistry at the University of Central Lancashire. Andrew previously worked at the University of Manchester and at Liverpool Hope University teaching in the life sciences. He also worked at St. Martin's College, Lancaster, where he contributed particularly to the development of learning through use of technology. His interests are in computer aided learning and their dimensional modelling.

Marion Nuttall is the Academic Enhancement Manager at the University of Cumbria and is an experienced lecturer and manager in higher education. She has taught and assessed students on a wide range of courses and has led professional development for staff on assessment, marking and moderating, and management of assessment at departmental level. She is particularly interested in students' engagement with assessment feedback to enhance their learning.

Acknowledgements

The authors would like to thank the following friends and colleagues for assistance in writing this book: Tim Barry, Penny Bradshaw, Matthew Brown, Christine Colcomb, Richard Lording, Caroline Marcangelo, Nicky Metcalfe-Meer, Kay Sambell and her colleagues at the Centre for Excellence in Teaching & Learning in Assessment for Learning at Northumbria University, Elliott Shaw and Bobbie Smith. In particular, we would like to thank all members of the Centre for the Development of Learning and Teaching at the University of Cumbria who gave us the space to write.



PART 1

INTRODUCTION AND CONTEXT



1 Introduction

The primacy of assessment

Research and experience tell us very forcefully about the importance of assessment in higher education. It shapes the experience of students and influences their behaviour more than the teaching they receive. The influence of assessment means that 'there is more leverage to improve teaching through changing assessment than there is in changing anything else' (Gibbs and Simpson 2004–5: 22). Tutors implicitly know the importance of assessment. Anecdotal experience tells us that, to a large extent, assessment activity in higher education is the learning activity. Students may take notes in lectures, seminars or from their reading, they may have been through the prescribed activities in laboratories or on field trips, but it is only when faced with assessment tasks that the majority seriously engage with that material. Tutors despair of trying to persuade students to undertake study which does not contribute in some way to their grades.

Sadly, though, university assessment practice lags well behind its equivalent in the school sector (Murphy 2006), relying largely on a limited range of tried (but not always tested) methods. It is dealt with in an *ad hoc* way (Swann and Ecclestone 1999a) and the situation is not mitigated by the 'amateur' status of many academics regarding assessment (Ramsden 2003: 177). We learn the craft of assessment informally through being assessed ourselves and through being part of a community of practice, but lack scholarship regarding assessment (Price 2005). Undoubtedly, most of us have survived this approach to professional learning reasonably unscathed but it is not a recipe for enhancement; it provides no reliable route for ensuring that research on assessment reaches those doing the assessing.

Assessment pressures and influences

The contemporary environment of higher education means that assessment cannot carry on unaltered; it is subject to too many pressures and influences which create a force for change. Increasing cohort size and the shrinking unit of resource creates pressure for more cost-effective assessment methods especially as assessment is very expensive and, in today's mass classrooms, can use more resources than teaching (Gibbs 2006b). This problem is

4 DEVELOPING EFFECTIVE ASSESSMENT IN HIGHER EDUCATION

exacerbated by **modularisation**, which has increased the volume of assessment as each small block of learning must be formally assessed and graded.

In addition, the student body is changing. Reliance on part-time work and other commitments appear to be turning students into very strategic learners (Kneale 1997) unwilling to devote effort to study which does not contribute to **summative assessment**. Tutors are increasingly teaching a much more diverse student body who challenge existing assumptions about what can be expected from new students (Northedge 2003a), with many non-traditional students needing greater support in making the transition to higher education. Poor early experience of assessment is associated with high student **attrition** rates (Krause 2001).

Moreover, the employability and graduate skills agenda is placing pressure on tutors to design assignments and examinations which assess a much broader range of achievement than in the past. Assessment is now expected to assess subject knowledge and a wide range of intellectual, professional and generic skills in a quality-assurance climate that stresses reliability with robust marking and **moderation** methods. Tutors are also facing pressure to modify assessment so that it supports learning through student involvement in assessment, prompt **feedback**, flexible and formative approaches and a wide variety of assessment methods.

In addition, assessment practices are being influenced by advances in technology. While computers afford the opportunity for online assessment, immediate feedback and computer-marked assignments, they also provide the breeding ground for the increase in **plagiarism**.

Within individual universities the mediation of regulations and the assessment process by departments, programme teams and individual tutors may be influenced, possibly constrained, by locally based, taken-for-granted assumptions, and even myths. Effective communication and academic development work may often be required to support programme teams in enhancing their assessment design and practice.

Finally, student evaluation through the National Student Survey (2006) has made student reactions to our programmes public for the first time, and assessment is proving to be the weakest area in the analysis. Competition in the new consumer market in higher education will mean that departments cannot neglect the student perspective for too long.

Policy

Perhaps the most obvious recent influence on assessment has been the policy climate in relation to quality assurance and enhancement. The quality assurance and accountability climate differs from nation to nation. In the UK,

institutional autonomy and self-regulation are now constrained by unambiguous public policy (Jackson 2000), largely in the guise of the 'academic infrastructure' of the Quality Assurance Agency (QAA 2006d). This includes a set of guidelines designed to create greater confidence in standards across British higher education, including the Framework for Higher Education Qualifications (**FHEQ**) which indicates the types of learning outcomes expected from different awards, **subject benchmark statements** for individual disciplines, various codes of practice, including one for assessment, and programme specifications. A key feature of external review of institutions (institutional audit) by the **QAA** is the extent to which the quality assurance procedures for any university comply with these guidelines.

At the heart of the QAA approach is the notion of constructive alignment between 'learning outcomes' and assessment. Assessment practice is judged primarily on whether it effectively measures the intended outcomes of a course of study in a valid, reliable and transparent way. This book acknowledges the centrality of this approach and its pervading influence on so much day-to-day institutional practice, and therefore an outcome-based method has been adopted throughout the text. However, such a philosophy is not accepted unquestioningly. Outcome-based course design represents a set of ideas which are currently fashionable in higher education quality assurance and educational development circles, but the approach is also open to criticism. Box 2.3 in Chapter 2 summarises the debate.

A further policy imperative emerged in the late 1990s in the UK. The Dearing Report (National Committee of Inquiry into Higher Education 1997), followed by various initiatives and the 2003 Higher Education White Paper (Department for Education and Skills 2003), placed considerable emphasis on raising standards of teaching and assessment in higher education. This included requirements for institutional learning and teaching strategies and strong encouragement for the professionalisation of academic staff in relation to learning, teaching and assessment. Some form of initial training for higher education lecturers is now widespread in British universities.

Research evidence

Publications now abound with tips for improving assessment and case study accounts of assessment practice. However, Knight and Yorke (2003: 209) argue that they largely represent a 'cottage industry' lacking a systematic theoretical basis for understanding judgements of achievement, and thus 'attempts to enhance assessment practices are built on sand'. This book attempts to distil the consistent elements of research findings to provide well-informed but intensely practical advice. In doing this, it is recognised that academics are by definition sceptical and will wish to see an

6 DEVELOPING EFFECTIVE ASSESSMENT IN HIGHER EDUCATION

acknowledgement of conflicting ideas and alternative perspectives, with any subsequent recommendations emerging from persuasive evidence.

Despite the evidence-based approach, we have attempted to write in an accessible way that does not require the reader to have prior knowledge of educational theory. *Leads into Literature* boxes will be used to provide routes into further reading or summarise areas of debate in relation to conflicting theories or controversial policies. In this manner, the book aims to provide strong guidelines explicitly supported by research.

Why another assessment book?

The dominance of assessment in the student experience and the social, economic and policy climate have led to a situation where assessment is in a state of flux, facing pressures for enhancement while simultaneously coping with demands to restrict its burden on students and staff. It is a demanding agenda but one which this book endeavours to embrace. The book recognises and welcomes the challenges presented above of assessment for learning, quality assurance, student numbers and diversity, modularisation, workload, plagiarism and technology. It also aims to provide a guide which focuses on all stages of the assessment cycle (see Figure 1.1). In this sense, the book is unique and comprehensive.

The book attempts to translate what is implied from research into the day-to-day demands of *doing* assessment in higher education. Our approach is informed by many years of experience struggling to improve assessment and use it creatively to influence students' learning. The poverty of assessment in higher education has made it tempting for assessment texts to advocate major institutional change; in our view this is an ideal rather than a realistic approach. Our experience and knowledge of the sector have persuaded us towards a more pragmatic approach recognising the limited appetite for change among academics facing huge pressures for productivity in other aspects of their role. Potential frustration for staff attempting change but constrained by institutional structures (Cuban 1988) is also acknowledged, so the book advocates practices which can have significant impact on the student experience yet have the potential to work within existing structures.

Thus, although we do not gainsay many of the conclusions of other scholars in the assessment field, they are not developed here. As Boud (2000: 159) suggests, 'one of the traps in arguing for a shift in assessment practice is to propose an unrealistic ideal that can never be attained'. In its place, we have attempted to write this guide within the bounds of what is possible in most university departments. The book focuses on discussion of issues, offering pragmatic solutions, and does not spend too much time advocating

the overhaul of a system which is too loosely coupled to be easily subject to change.

Having said that, the text does alert staff, particularly those in positions of responsibility, to some of the weaknesses in existing assessment infrastructure such as modularisation, constraining regulations and the consequences of tutor discretion in **examination boards**. Overall, the book aims to combine a clear academic rationale for good practice with concrete advice and living examples of successful assessment interventions.

Situated practice

Tempting as it is, educational research and theory do not translate simply into ideas for educational practice. Laurillard (2002: 62–3) emphasises the ‘situated character of all learning’ and the impossibility of defining ‘reliable prescriptions for teaching strategies’. Readers’ knowledge of assessment is situated in the context of their own experience and in the particular traditions, expectations and needs of different academic subjects. What we offer is based on our own experience and interpretation of the literature, but every recommendation and activity has to be adapted to the reader’s local context. This includes tutors, their skills, experience, time, enthusiasm and interests. It also includes students, their previous education, backgrounds, knowledge, skills, level of study and motivation. Finally, it also includes contextual issues such as group size, resources, regulations, and disciplinary and professional requirements.

Thus, although the growing evidence base of research on assessment provides a useful basis on which to build and review practice, it leaves the onus on tutors and teaching teams to develop and critically evaluate assessment processes and procedures as they are used and developed within their local context. Thus, while the evidence base can inform institutional and departmental policy, it will require mediation to suit local contexts and students groups.

Overall, we need to listen to Black and Wiliam (1998b) when they assert that there is no quick fix which will give rapid rewards in relation to assessment. Earl (2003), drawing on the work of Cuban (1988), refers to the notion of first-order and second-order change. First-order change is making existing procedures more efficient and effective, and we have all seen institutional strategies designed with this end in mind – speeding up the time for return of marked assignments, improving assignment feedback proformas, streamlining procedures for assessing claims for extenuating circumstances, and introducing plagiarism-detection software to improve **malpractice** procedures. This text does provide advice on these procedural matters but, as Crook *et al.* (2006) point out, equitable and consistent procedures are not sufficient

8 DEVELOPING EFFECTIVE ASSESSMENT IN HIGHER EDUCATION

to deliver good-quality assessment practice. Procedural changes and efficiencies will struggle to determine individual practices. For example, procedures may ensure assignment briefs are published four weeks in advance of the deadline and always include **assessment criteria**, but they do not ensure the appropriateness of the **assignment** or the quality of the criteria. Such first-order changes do not get to the heart of individual practice. That is second-order change (Earl 2003): change designed to alter the fundamental ways staff operate. It is the latter that we aim for in writing this book. We are hoping to encourage a transformative approach to thinking about the purposes of assessment, enabling staff to comfortably '[accept] and [embrace] the subjectivity of judgement' (Clegg and Bryan 2006: 224) so that they, in part, are liberated to review the contribution of their practice to student learning.

Audience

This book is aimed at lecturers in higher education and others with responsibility for the assessment of taught programmes of study. While we hope it will be of particular use to new lecturers, we have also written it with more experienced staff in mind: those embarking on a new role or responsibility in relation to assessment, writing a new module or taking on programme leadership. It is not aimed at educational scholars, for whom there are more research-focused texts available, although they may choose to follow up the leads into literature offered within the chapters. We particularly commend the book to programme leaders, whom we see as the vital link in the chain between individual tutor intentions and the likelihood of providing a coherent assessment experience for individual students.

In addition, the book is also aimed at those with subject management, quality assurance and educational development remits who are seeking information regarding assessment strategy and management.

Plan of the book

Figure 1.1 sets out the structure of the book. Above all, the book is distinguished from its predecessors by attempting to capture all the stages of assessment from initial task design to final examination board and evaluation. These stages are reflected in the central spine of the diagram (rectangular boxes). One might characterise these chapters as dealing with the day-to-day practice of assessment. The oval or round shapes characterise elements of assessment research and practice which you may wish to consider in developing your assessment strategy. The two diamond shapes indicate processes which influence all stages of the assessment cycle: the management of

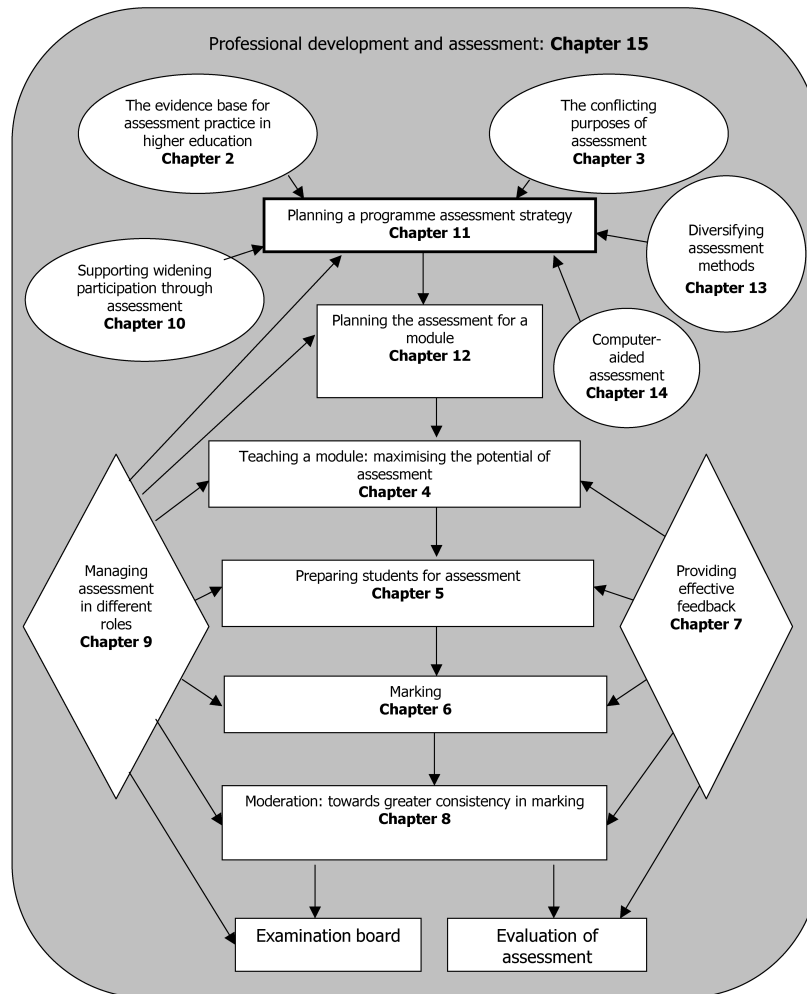


Figure 1.1 Structure of the book

assessment in different roles and providing effective feedback. Finally, successful development of assessment relies on the active engagement and learning of tutors, and therefore Chapter 15 is shown as underpinning all other chapters of the book.

The book falls naturally into three parts. Part 1 summarises research on the relationship between assessment and learning and critically explores the difficulty of reconciling the various principles underlying assessment with its different purposes. It provides both a challenge to current assessment practice

10 DEVELOPING EFFECTIVE ASSESSMENT IN HIGHER EDUCATION

and outlines the research and theory which underpin the advice contained in subsequent chapters. Part 2 focuses on the day-to-day matters associated with dealing with assessment, including recognition of the new emphasis on assessment and feedback which 'promotes learning' (QAA 2006c), and working with the needs of diverse students, including those with disabilities. Part 3 is aimed at the design stage of the assessment cycle. It stresses the importance of taking a programme-level approach to developing assessment, as much as anything to ensure a balance of practice which can meet the often conflicting demands on our assignments and examinations. It provides ideas for broadening the range of assessment to meet its different purposes, including an overview of online assessment. Finally, we examine the issues and good practice associated with developing tutors and teaching teams with respect to assessment.

The book takes a broad view of the purposes of assessment, including recognising the pressures for quality assurance and standards discussed earlier. The implications of these policies, including the QAA code of practice on 'Assessment of Students' (QAA 2006c), are woven throughout the text, and Box 1.1 identifies where specific information can be found in the book.

How to use this book

The book is not intended to be read in a linear fashion but is designed to help individual tutors and teaching teams with relevant advice depending what point of the assessment cycle they wish to examine. Some suggestions for how to use the book are as follows:

- New tutors or postgraduate teaching assistants may wish to start with Chapter 4 which deals with the 'assessment basics' needed when beginning to teach a module for the first time. Chapter 6 also provides advice on marking.
- Tutors could use the bulleted lists which appear in most chapters as checklists, for example to ensure module outlines contain appropriate assessment information, to check they are aware of assessment responsibilities, or to test the impact of assessment practice on international students.
- Programme leaders or heads of department could use short sections of the book as pre-reading for team meetings. For example, the chapter on providing effective feedback (Chapter 7) or the section on designing assessment to reduce plagiarism (Chapter 4) might be used to stimulate discussion about departmental practice.
- Programme leaders could draw on Chapter 11 at the beginning of a course design process, using the ideas presented to interrogate their

Box 1.1 Code of Practice

The text below sets out Appendix 1 of the *Code of Practice for the Assurance of Academic Quality and Standards in Higher Education*, Section 6: Assessment of Students (© The Quality Assurance Agency for Higher Education 2006) and identifies which chapters address each specific principle.

The Precepts**General principles**

1. As bodies responsible for the academic standards of awards made in their name, institutions have effective procedures for:

- i designing, approving, monitoring and reviewing the assessment strategies for programmes and awards

Chapter 11

- ii implementing rigorous assessment policies and practices that ensure the standard for each award and award element is set and maintained at the appropriate level, and that student performance is properly judged against this

Chapter 12

- iii evaluating how academic standards are maintained through assessment practice that also encourages effective learning.

Chapter 9

2. Institutions publicise and implement principles and procedures for, and processes of, assessment that are explicit, valid and reliable.

Chapters 2 and 4

Contribution to student learning

3. Institutions encourage assessment practice that promotes effective learning.

Chapters 2, 4, 7, 10, 11, 12, 13, 14.

Assessment panels and examination boards

4. Institutions publicise and implement effective, clear and consistent policies for the membership, procedures, powers and accountability of assessment panels and boards of examiners.

Chapters 8 and 9

Conduct of assessment

5. Institutions ensure that assessment is conducted with rigour, probity and fairness and with due regard for security.

Chapters 6, 8 and 9

Amount and timing of assessment

6. Institutions ensure that the amount and timing of assessment enables effective and appropriate measurement of students' achievement of intended learning outcomes.

Chapters 4, 11 and 12

12 DEVELOPING EFFECTIVE ASSESSMENT IN HIGHER EDUCATION

Marking and grading

7. Institutions have transparent and fair mechanisms for marking and for moderating marks.

Chapters 6 and 8

8. Institutions publicise and implement clear rules and regulations for progressing from one stage of a programme to another and for qualifying for an award.

Chapter 9

Feedback to students on their performance

9. Institutions provide appropriate and timely feedback to students on assessed work in a way that promotes learning and facilitates improvement but does not increase the burden of assessment.

Chapter 7

Staff development and training

10. Institutions ensure that everyone involved in the assessment of students is competent to undertake their roles and responsibilities.

Chapters 9 and 15

Language of study and assessment

11. The languages used in teaching and assessment are normally the same. If, for any reason, this is not possible, institutions ensure that their academic standards are not consequently put at risk.

Professional, statutory and regulatory bodies' requirements

12. Institutions provide clear information to staff and students about specific assessment outcomes or other criteria that must be met to fulfil the requirements of PSRBs.

Chapter 11

Assessment regulations

13. Institutions review and amend assessment regulations periodically, as appropriate, to assure themselves that the regulations remain fit for purpose.

Chapter 9

Student conduct in assessment

14. Institutions encourage students to adopt good academic conduct in respect of assessment and seek to ensure they are aware of their responsibilities.

Chapters 3, 4, 5 and 14

Recording, documenting and communicating assessment decisions

15. Institutions ensure that assessment decisions are recorded and documented accurately and systematically and that the decisions of relevant assessment panels and examination boards are communicated as quickly as possible.

Chapter 9

- programme assessment strategy. Module leaders are encouraged to read Chapter 12 as a first step in rethinking the assessment of their modules.
- Staff developers could use the text as a resource in designing staff development workshops, for example by using the case study approach in Chapter 3 to analyse how well diverse types of assignment provide for the different principles of assessment or using Chapter 6 as the pre-reading for a workshop on marking for post-graduate teaching assistants.
 - Practitioner researchers could use the references in the *Leads into Literature* boxes as a stimulus for further investigation of aspects of assessment.
 - Quality assurance teams could use Chapter 11 in the training for validation or accreditation panels. It can help them identify the questions they might want to ask in testing the merit of a new programme assessment strategy.

Cross-references are used throughout the book to assist readers in finding broader information of relevance to the topic of a particular chapter.

A note about terminology

Various different terms are used to refer to the same entity in higher education across English-speaking countries. Therefore, in order to avoid considerable confusion and repetition in the book, we have adopted certain terms as follows:

Assessment task – any item of assessment whether examination, test, coursework or direct observation.

Assignment – coursework usually undertaken by a student or students in their own time and not under controlled conditions.

Examination – an assessment task undertaken under controlled conditions.

Test – an assessment task taken in semi-controlled conditions such as an in-class or online test, usually of a relatively short duration.

Assessment strategy – the procedures adopted to assess student learning in a given module or programme.

Module – a specific unit of study or block of learning which is separately assessed. Combinations of modules form a programme of study.

Programme – the overall curriculum followed by an individual student, normally comprising a specified set of modules or option choices.

Course – unlike programme and module, which are used very specifically, the term *course* is used generally, to refer to any organised scheme of teaching.

14 DEVELOPING EFFECTIVE ASSESSMENT IN HIGHER EDUCATION

Curriculum – like ‘course’, used generally to refer to all aspects of the student learning experience. It includes both the syllabus (content) and the teaching, learning and assessment methods.

Year – many staff in higher education have replaced the term ‘year’ with ‘level’ to represent the stage of learning, because the diversity in modes of study means that students are often spending more than a year on a programme level. Thus, level 1 is equal to year 1 of a full-time undergraduate programme. We have chosen to use ‘year’ as the indicator of level of study because it is readily understood and because various labels exist for different levels. For example, level 1 is also referred to as foundation or level 4 depending on the framework in use. The use of the term ‘year’ implies no assumption that all students are full-time. (See Box 11.3 for a discussion of levels.)

Attribution/attributable – we have selected these terms to refer to the extent to which an assignment can be reliably attributed as the work of an individual student. The word ‘authenticity’ is frequently used in this way, but we have rejected that term because it is also commonly used to mean an assignment which mirrors realistic demands outside the university. Using the term for both meanings would be confusing.

A full glossary of terms and acronyms used in this text is set out in the Appendix.

Conclusion

This text is offered as a comprehensive resource based on research, public policy and experience. As with most things educational, there are no right or simple answers that can be employed across the messy business of providing programmes of learning for adults. There are only more or less likely solutions to problems and they will be affected by you, your students, the learning environment, and the **subject discipline** in many different ways. The temptation might be to hold back from advice in such unpredictable circumstances, yet that is a recipe for leaving things as they are in an environment which is hardly static. The pressures discussed in the opening paragraphs emphasise the importance of taking action, and that action should at least be as well informed as possible. Having said that, we have had to be fairly sparing in our use of evidence in order to balance background information with practical advice. We realise it is presumptuous to claim this as a guide to good practice, and indeed we would prefer our advice to provoke you, even enrage you, rather than leave you untouched. If we wish for anything, it is that tutors and teaching teams seriously engage in debate about assessment and decide for themselves what constitutes good practice.

2 The evidence base for assessment practice in higher education

There is a growing body of research into higher education assessment on which we can begin to build robust policy and practice decisions. This book has drawn on that evidence as a basis for its advice on 'good practice'. Therefore, in advance of our discussions on dealing with and developing assessment in Parts 2 and 3, respectively, this chapter summarises what the research evidence tells us about assessment practice in higher education. Of course, a methodical appraisal of the field would constitute a substantive book in its own right, whereas for the purposes of this text we shall limit ourselves to summarising the main conclusions of research in an accessible way. The *Leads into Literature* boxes will direct readers seeking more information to appropriate research literature. Box 2.1 introduces four useful systematic reviews of the assessment literature.

Assessment purposes have been categorised usefully as assessment *of* learning, assessment *for* learning and assessment *as* learning (Earl 2003). Assessment *of* learning characterises how we may traditionally view assessment. It involves making judgements about students' summative achievement for purposes of selection and certification, and it also acts as a focus for institutional accountability and quality assurance – for example, the number of 'good' degrees awarded is used as a key variable in university league tables. On the other hand, assessment *for* learning is **formative** and **diagnostic**. It provides information about student achievement which allows teaching and learning activities to be changed in response to the needs of the learner and recognises the huge benefit that feedback can have on learning (Black and Wiliam 1998a). Finally, assessment *as* learning can be defined in two inter-linked ways. First, at a very straightforward level, tackling assignments and **revision** is when higher education students *do* much of their learning. Second, assessment *as* learning is a subset of assessment *for* learning and sees student involvement in assessment, using feedback, participating in peer assessment, and self-monitoring of progress as moments of learning in themselves (Black and Wiliam 1998a). Students come to have a better understanding of the subject matter and their own learning through their close involvement with assessment.

This summary of relevant literature focuses primarily on what the

Box 2.1 Leads into literature: systematic reviews of assessment literature in higher education.

Gibbs and Simpson (2004–5) have conducted a thorough review of research studies which they have drawn upon to develop a set of eleven ‘conditions under which assessment supports learning’ which are offered as a useful reference point for developing assessment practice.

An earlier comprehensive review of formative assessment by Black and Wiliam (1998a) includes research from both school and university settings with consistent findings regarding the significance of feedback on learning compared with other elements of teaching, and the distinctive benefits for learning accruing from peer and self-assessment.

Struyven *et al.* (2002) have also conducted a useful review of the research into students’ perceptions of assessment in higher education although they conclude that the literature and research on this area are relatively limited, particularly in relation to students’ perceptions of specific types of assessment.

Finally, Elton and Johnston (2002) have drawn on a wide range of studies in their critical review of research on assessment in higher education, with particular emphasis on challenging some of the underlying assumptions in our assessment traditions.

These publications will provide the reader with extensive lists of further reading across the range of assessment topics.

literature tells us about assessment *for* learning and *as* learning, but will also briefly address assessment *of* learning. Chapter 3, which debates the major tensions in contemporary assessment practice, will discuss assessment *of* learning in more detail.

The relationship between assessment and learning

The assessment strategy of a particular course has a major impact on student activity (Snyder 1971). It influences the approach students adopt towards their learning, how much time they spend on their studies, how widely they study the curriculum, and whether they grasp the key concepts of the subject. There is also evidence of a significant, negative ‘backwash’ effect (Biggs 2003) on student learning and achievement from poorly conceived assessment strategies. Therefore, tutors who neglect to pay attention to their assessment practices are ignoring an important opportunity to enhance students’ effort, approach and outcomes.

Approaches to learning

The concept of student approaches to learning has developed from the initial work of Marton (1976), with findings replicated in many other studies using a phenomenographic approach. Prosser and Trigwell (1999: 88) provide an excellent overview of the history of the seminal studies in this field, concluding that the 'approaches to study' work has 'resulted in an approach to research in student learning which has played a major role in bridging the gap between educational research and the practice of university learning and teaching'.

In the research, two approaches to learning by students which consistently emerge are a **surface** approach and a **deep** approach (Marton and Saljo 1997; Ramsden 2003). The student's conception of learning and their *intention* when studying are central to the approach they take. Students adopt a surface approach when their intention is to cope with the requirements of the task but with little personal engagement or aim to understand the material. Such students want to get by with minimum effort and tend to focus on the detail of the knowledge, memorising the information or procedures, for example rote learning for an examination. As a result, students do not grasp the overall meaning of their studies, develop limited conceptual understanding of the material and have poor-quality learning outcomes (Entwistle 1997).

In contrast, students who adopt a deep approach aim to understand ideas and are intrinsically interested in their studies. The learning strategies they use include relating information and ideas together and to their own experience and looking for patterns, principles and meaning in the texts. This approach leads to higher-quality learning outcomes for the student.

An approach to learning is not a fixed characteristic of an individual but is influenced by their perception of the learning environment, most particularly the assessment task (Morgan and Beatty 1997; Biggs 2003). Appropriate assessment can encourage students to adopt a deep approach to learning, and the contrary is true for poorly designed assessment. If students perceive that a task requires memorisation and reproduction of facts, then that is what they will do. The link between assessment method and student approach to learning is vital for the design of assessment in higher education, and later chapters on teaching a module (Chapter 4) and developing programme and module assessment strategies (Chapters 11 and 12) will discuss the practical implications.

Biggs and Moore (1993) have argued that a range of course characteristics encourage students to take a deep approach. These include a clear structure to the knowledge base of the course, so that the content is integrated and new topics relate to previous knowledge. Moreover, courses should encourage students' intrinsic interest in the subject matter. Finally, they should involve

18 DEVELOPING EFFECTIVE ASSESSMENT IN HIGHER EDUCATION

learner activity and interaction with others. Other features in the context are also important, such as using ‘teaching and assessment methods that foster active and long term engagement with learning tasks [and] . . . opportunities to exercise responsible choice in the method and content of study’ (Ramsden 2003: 80). These criteria are helpful in thinking through the design of assessments.

In summary, the research evidence suggests that if the nature of the learning context is changed, and assessment is the most influential element of that context (Elton and Johnston 2002), there is a likelihood that students’ approach to learning will change (Prosser and Trigwell 1999). Chapters 11–13 will concentrate on how programme and module design can use this pattern of student response to develop effective programme and module assessment strategies.

Students’ perceptions of assessment

Changing student approaches is a complex business, with evidence from some studies that it appears relatively easy to induce a surface approach (Struyven *et al.* 2002) but less straightforward to encourage a deep approach. A key explanation for this difficulty appears to lie in the research on students’ conceptions of learning and their perceptions of assessment. Prosser and Trigwell (1999) draw on research studies to show how student learning is a result of an interaction between a student and the learning situation and this is unique for every student in every different learning context. It will vary between modules because aspects of the situation will differ. Various factors contribute to this, including students’ prior experience of education, their perception of the current situation and their approach to learning. In relation to assessment, the students’ perception of what the assessment requires affects the approach they take (Prosser and Trigwell 1999). Students behave differently because they perceive tasks differently. Changing the assessment may change the approach of some students who perceive the new requirements appropriately, but will not necessarily change every student’s approach to learning. As Ramsden (2003: 66) says of a group of students who did not change in response to assessment encouraging a deep approach: such students ‘have brought with them a predisposition to use a surface approach which they had previously developed in response to similar situations. Like all of us, they carried their history of learning along with them’. Entwistle and Tait (1990) discovered a relationship between students’ approaches to study and their assessment preferences, with students who take a surface approach preferring teaching and assessment procedures that supported that approach and vice-versa. These habitual tendencies (Ramsden 2003) may work against students engaging in effective learning in higher education. The implications for preparing students for assessment will be discussed in Chapter 5.

While students' 'approaches to learning' and 'conceptions of learning' are key factors in understanding the relationship between assessment and learning, research has also identified a range of related features of assessment which impact on student behaviour and the student learning experience, and these are discussed below.

Strategic approaches and cue seeking

A feature of modern modular course structures means that most assessment has a summative function, and this places a pressure on students throughout their programmes to focus on assessment rather than learning (Heywood 2000). Evidence suggests that students are increasingly taking a strategic approach to their studies, focusing their effort on assessment-related tasks (Kneale 1997; Gibbs 2006b). Students become 'cue-conscious' (Miller and Parlett 1974), concentrating on passing the assessment. The latter activity may, unintentionally, be at the expense of understanding the subject matter (Gibbs 2006a). This cue-seeking behaviour, where students are determined to work out what the tutor is looking for rather than developing an understanding of the subject matter of the assignment, has been identified in other studies (Bloxham and West 2007). As Ramsden (2003: 67) asserts, much of student activity is about adjusting to the perceived requirements of the tutor – the 'hidden curriculum'. Research suggests that students who fail to pick up on such cues are not likely to do well (Miller and Parlett 1974). This has major implications for the choice of assessment and the guidance provided for students in order to direct their efforts towards appropriate activity: 'unless assessment tasks mirror the official curriculum, they will erode it. Assessing outside, or below, the curriculum gives irrelevant or counter-productive tasks a false value' (Biggs, 2003: 63).

Time devoted to studying

Different assessment regimes are also related to different amounts of time devoted by students to their out-of-class studies. Gibbs and Simpson (2004–5) draw on a range of research to show that time spent studying increases achievement, although it can be wasted by ineffective surface learning. In addition, more frequent assessment tasks are associated with greater time allocated to study. Moreover, there is evidence that students will work harder in preparation for some modes of assessment than for others. Traub and MacRury (1990), in a review of research on multiple choice and free response tests over the previous 20 years, found that students appear to prepare better for free response (that is, where the student must construct the answer rather than select from a given set of answers).

In addition to assessment influencing the amount of time spent studying,

20 DEVELOPING EFFECTIVE ASSESSMENT IN HIGHER EDUCATION

it can also affect when students study. For example, infrequent examinations and coursework tasks encourage students to bunch all their **learning hours** together in the time immediately preceding the examination or submission date (Gibbs and Simpson 2004–5). Overloading of students through excessive amounts of content is also associated with a surface approach to learning (Ramsden 2003).

Student reactions to assessment

Most students undergo some stress in relation to assessment, but few suffer from severe psychological difficulty (Heywood 2000). Indeed, as Heywood (2000: 149) asserts, a level of stress ‘would appear to be essential for learning’. Nevertheless anxiety provoked by assessment can encourage students to adopt surface approaches to learning (Rust 2002) and make them less likely to take a deep approach (Fransson 1977). There is particular evidence of stress associated with examinations (Falchikov 2005), with students generally preferring coursework (Gibbs and Simpson 2004–5), although there is support for the view that size and volume of tasks can act as stress factors (Sarros and Densten 1989). Heywood (2000) argues that preparation for assessment is a necessary response to reducing anxiety, supported by research showing that improving students’ understanding of goals and standards is associated with enhanced achievement (Rust *et al.* 2003), particularly with unfamiliar forms of assessment (Hounsell *et al.* 2006). Chapter 5 provides practical ideas to help students understand the criteria and standards of their assignments and examinations.

The relationship between feedback and learning

The most important aspect of the assessment process in raising achievement is the provision of feedback (Black and Wiliam 1998a; Gibbs and Simpson 2004–5). Research indicates that students value feedback (Hartley *et al.* 2002; Weaver 2006) despite anecdotal staff views and contrary research evidence regarding how students fail to engage with it (Gibbs and Simpson 2004–5). However, not all feedback is perceived as useful by students (Black and Wiliam 1998a; Hounsell *et al.* 2006), and concerns over feedback have been strongly reflected in the first two years of the UK National Student Survey (2006). Entwistle *et al.* (1989), studying engineering students, showed that early failure was related to students gaining no feedback at all in their first term. According to Hounsell (2003), this combination of students choosing not to or being unable to use feedback, and staff cynicism that their efforts are wasted, creates a vicious ‘downward spiral’ in relation to the potential gains from feedback.

Feedback has little value unless it is timely and students pay attention to it, understand it, and act on it (Gibbs and Simpson 2004–5), and various studies suggest or investigate practical activities to help students engage with it (Nicol and Macfarlane-Dick 2004). Black and Wiliam (1998a), in an extensive review of literature on formative assessment, concluded that feedback in the form of comments can have a significantly greater effect on future improvement than feedback that is limited to a grade or mark. Knight and Yorke (2003) argue that feedback is mostly likely to be useful to learners if they are willing and able to expose their areas of weakness and confusion with a topic. This is supported by Black *et al.*'s (2003) work on formative assessment with school teachers, where students found that revealing their problems was worthwhile and led to gaining help. Reflective assignments such as learning journals attempt to tackle this dilemma by providing students with an opportunity to reflect on their strengths and weaknesses as learners without losing marks, but these provide their own problems for assessment (Gibbs, 1995; see also Chapter 13). Recent studies have placed greater importance on the notion of **feed forward** (Torrance 1993; Hounsell 2006), which focuses on what a student should pay attention to in future assessment tasks, and ensuring that feedback is embedded in day-to-day learning activities as well as provided in response to formal assignments, for example as in-class and online activities (Laurillard 2002).

Using feedback to adjust teaching

It is not just students who need to act on feedback. For assessment to function in a formative way that supports students' future learning, the findings have to be used to adjust teaching (Black and Wiliam 1998a; Prosser and Trigwell 1999; Nicol and Macfarlane-Dick 2006). Difficulties with a particular concept or problem may signal that further or different tuition is needed. Angelo and Cross (1993) and Nicol and MacFarlane-Dick (2004) provide a range of 'classroom assessment techniques' designed to assist staff in gaining immediate feedback from students which can be used to revise teaching strategies. However, course structures with short modules can make it difficult for individual tutors to respond to the information about student learning emerging from summative assessment. Chapter 4 on 'teaching a module' will provide practical advice on embedding formative assessment in your teaching and being responsive to what you find out.

Students as assessors

Recent work in the field of feedback is focusing on the importance of the student as self-assessor: someone who is able to provide their own feedback because they understand the standard they are aiming for and can judge and

22 DEVELOPING EFFECTIVE ASSESSMENT IN HIGHER EDUCATION

change their own performance (that is, self-regulate) in relation to that standard (Nicol and Macfarlane-Dick 2006). This is assessment *as* learning (Klenowski 2002; Earl 2003) and is firmly located in Sadler's (1989) view that improvement involves three key elements:

- students must know what the standard or goal is that they are trying to achieve (assessment guidance);
- they should know how their current achievement compares to those goals (feedback);
- they must take action to reduce the gap between the first two (applying feedback to future assignments).

As Black and Wiliam (1998a: 15) assert, 'self-assessment is a sine qua non for effective learning', and certainly systematic reviews of research (Black and Wiliam 1998a; Falchikov 2005) indicate strong positive benefits to students of being involved in their own assessment.

If students are to become specialists within a subject discipline, they need to develop the capacity to assess quality in that field. Involving students in assessment provides an authentic opportunity for them to learn what 'quality' is in a given context and apply that judgement to their own work (Black *et al.* 2003). The context might be solving a problem, doing an experiment, creating a design, or writing an essay. Thereby the student becomes aware of what the goals or standards of the subject are (Earl 2003), a precondition of taking responsibility for their work (Swann and Ecclestone 1999a). Feedback allows the student to see their performance against those goals. This view is supported by Black and Wiliam (1998a) when they stress that peer and self-assessment are the key to learning from formative assessment. It is not enough for a tutor to tell a student what they need to do to improve ('your writing is too descriptive', 'you have mistaken correlation for cause') if the student does not understand what these comments mean in relation to the subject or their writing. They cannot do anything about it until they begin to share the tutor's conception of the subject (Sadler 1989). Box 2.2 develops this topic in relation to peer and self-assessment.

The assessment *as* learning approach is challenging prior ideas about the separation of formative assessment (assessment *for* learning) and summative assessment (assessment *of* learning) (Carless *et al.* 2006; Hounsell 2006) and replacing it with the notion of 'learning-oriented assessment' characterised as 'when tasks are "fit for purpose"; when students are involved in the assessment process in ways which support the development of evaluative expertise; and when feedback is forward-looking and can be acted upon' (Carless *et al.* 2006: 396). It is argued that assessment and feedback activity of this nature does not just contribute to learning at university but develops learning and evaluative skills essential for employment and lifelong learning (Boud and

Box 2.2 Leads into literature: peer and self-assessment

Peer and self-assessment involve students in assessing themselves and other students and are widespread in higher education today, often linked to wider innovation in learning and teaching (Stefani 1998). Numerous studies of peer assessment exist with considerable agreement about the benefits (Bostock 2000; Topping 2000; Falchikov 2005), including a greater sense of accountability, motivation and responsibility, and an increase in the speed of feedback (Black *et al.* 2003). In particular, peer assessment increases understanding of the subject matter, standards required and students' own achievement; it involves using disciplinary knowledge and skills in order to make judgements (Bostock 2000).

Peer and self-assessment are also seen as valuable in helping students develop important skills for lifelong learning (Authur 1995; Boud 2000), such as self-evaluation, giving feedback, justifying a point of view and negotiation skills. 'If assessment processes are intended to enhance student learning then it follows that students must be enabled to reflect on their current attainment' (Stefani 1998: 346). Black *et al.* (2003), working with secondary school teachers, found that peer assessment helped students develop the objectivity required for self-assessment and thus the capacity to direct their own work and thereby become autonomous learners. They conclude that peer and self-assessment 'make unique contributions to the development of students' learning – they secure aims that cannot be achieved in any other way' (2003: 53).

Concerns about peer assessment, such as imprecise marking, are not supported by the research, which generally finds good levels of agreement between staff and students where students are working with criteria (Falchikov and Goldfinch 2000) and have been trained (Falchikov 2005). Various other strategies have been successfully developed to address student and staff concerns, such as use of **double anonymous** peer marking (Bostock 2000). Falchikov (2005) presents evidence that the greatest degree of agreement between student and tutor scores in self-assessment occurred with more detailed grading scales. In general, research on peer assessment has tended to focus on agreement between tutor and student grades, whereas if it is used at the formative stage with the emphasis on feedback, many of the worries regarding grading can be discounted (Falchikov 2005). In addition, feedback from students may be superior in some ways to teacher feedback. Black *et al.* (2003: 77) found that peer feedback is not as 'emotionally loaded', students will accept criticism more readily from peers, and the language used by peers may be easier for students to understand (Bloxham and West 2004).

Studies consistently report positive responses to peer marking from students (Bostock 2000; Orsmond *et al.* 2000; Black *et al.* 2003) who claim it has made them think more, become more critical, learn more and gain in confidence. This positive response is reflected in Falchikov's (2005) substantive summary of research, although she notes some areas of difficulty and the improvement that comes from familiarity with the method. Even where students were negative about the

24 DEVELOPING EFFECTIVE ASSESSMENT IN HIGHER EDUCATION

experience, researchers continued to find evidence of benefits in terms of motivation and critical thinking (Oliver and Omari 1999).

In summary, peer and self-assessment have emerged as important tools in the tutor's repertoire for their many potential benefits. For practical advice on implementing peer assessment, including training students, see Chapter four, and for advice on peer marking of group assessments, see Chapter 7.

Falchikov 2006), characterised as 'sustainable assessment' (Boud 2000; Hounsell 2006).

While the latter ideas are in early stages of development in terms of practice-based research, they do underpin the growing acknowledgement of the importance of involving students in assessment. This attention has stimulated a range of empirical and theoretical studies, practical projects, innovations and action research which are used extensively in the following chapters of this book.

The validity of assessment tasks

While much of the research evidence discussed earlier has focused on assessment *for* and *as* learning, there is extensive research on the effectiveness of assessment practices as measurement *of* learning, including a focus on the validity of assessment tools. Here validity means that assessment tasks are assessing the stated learning outcomes. Prosser and Trigwell (1999) point out that assessment does not always test what we think it does and sometimes cannot reveal the qualitative differences in understanding between different students. Knight and Yorke (2003) argue that assessing higher-order learning in any discipline is not uncomplicated and judging the products (essay, exam script) of student learning has its limitations. For example, Entwistle and Entwistle (1997) show that where students are able to reproduce in their examination answer the structure of the topic as given by the tutor, they can give the impression of well-structured understanding. Similarly, Knight (2000) found that if a student has been given considerable support and direction, they may produce an assignment of similar quality to one produced in another context where the questions are not closely aligned to the teaching and the student has to work unsupported. Although the products look the same, they do not represent the same achievement.

In other words, our assessment tasks may be assessing learning at a lower level than that intended. In order to develop this idea, we need to think about what we mean by different 'levels' of achievement.

A confusing number of taxonomies or frameworks of thinking have been developed to assist instructional design (Moseley *et al.* 2005). Brown *et al.*

(1997: 38) argue that while not all lecturers may find such taxonomies suitable for their programmes, 'a classification of the kinds of skills and capabilities that one wants students to develop is a necessary first step in developing an assessment system'. Biggs and Collis's (1982) structure of observed learning outcomes (SOLO) taxonomy (see Figure 2.1) provides such a framework to discriminate between different stages of achievement. The learning is described not just in relation to the content the students should learn, but also what they are able to do with the knowledge. Thus, in the SOLO taxonomy, the selection of appropriate verbs to describe student capabilities is fundamental.

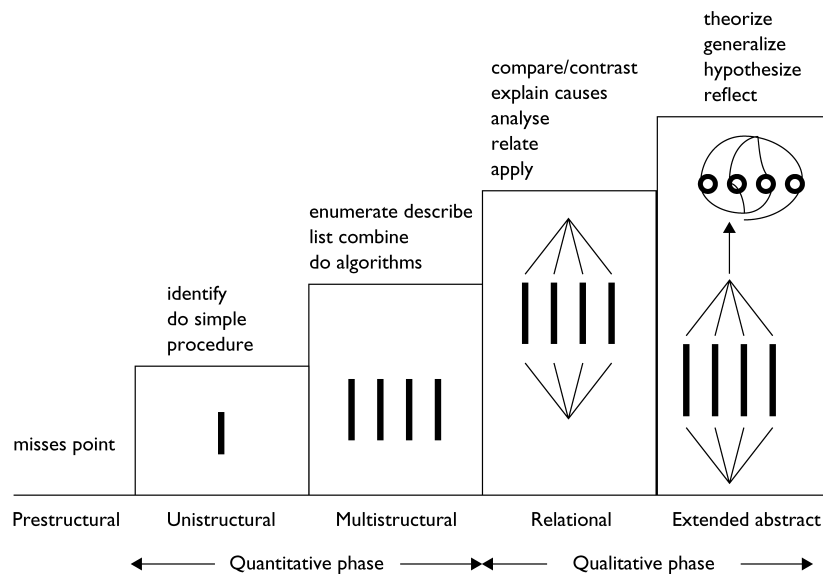


Figure 2.1 A hierarchy of verbs that may be used to form curriculum objectives

Source: Biggs (2003: 48).

This is not dissimilar to Anderson and Krathwohl's (2001) approach in revising Bloom's taxonomy of educational objectives. However, their taxonomy (Table 2.1) has two dimensions, the *knowledge* dimension and the *cognitive process* dimension. In a similar way to the SOLO taxonomy, the *cognitive process* dimension enables the tutor to identify an appropriate verb which should be used to express the learning outcome. The other dimension determines what knowledge (the noun) the verb is describing, and delineates between the facts a student needs to be familiar with the discipline; conceptual knowledge such as knowledge of classifications, principles, theories, models and structures; procedural knowledge, that is, knowing how to do

26 DEVELOPING EFFECTIVE ASSESSMENT IN HIGHER EDUCATION

something, including techniques, skills and methods of inquiry; and meta-cognitive knowledge, knowledge of self and cognitive tasks, and methods of learning and organising ideas (Anderson 2003).

Table 2.1 The taxonomy table (Anderson and Krathwohl, 2001)

<i>Knowledge dimension</i>	<i>Cognitive process dimension</i>					
	<i>1. Remember</i>	<i>2. Understand</i>	<i>3. Apply</i>	<i>4. Analyse</i>	<i>5. Evaluate</i>	<i>6. Create</i>
A. Factual	X				Z	
B. Conceptual		X			Z	
C. Procedural			Y			
D. Metacognitive						

Source: adapted from Anderson (2003: 29).

Anderson (2003) argues that this taxonomy table helps teachers design appropriate assessment because it enables them to work out prototypical ways of assessing objectives that fall within the relevant cells. For example, it is easy to see that a multiple-choice exam could assess memory of factual knowledge or possibly understanding of conceptual knowledge (the cells marked X in Table 2.1). However, application of procedural knowledge (cell marked Y) will need an assessment task, for example problem solving or case study analysis, which requires students to demonstrate not just that they can remember or understand something, but that they can use it.

Such a taxonomy is helpful in thinking about what different assessment tasks are testing. Here are two religious studies essays:

- Outline one of the five pillars of faith and explain why it is important for Muslims (Year 1)
- Critically assess the claim that British society has not adequately provided for the needs of its Muslim population (Year 3)

The first example appears to demand recall of factual information and understanding of conceptual knowledge, again in the cells marked X in Table 2.1. The second essay appears to be demanding 'evaluation', the cells marked Z in Table 2.1, a relatively high-level cognitive skill requiring good command of the subject matter. Unfortunately, the questions alone are insufficient to determine whether they measure qualitatively different achievements. For example, if the issues in the year 3 question have been carefully rehearsed in a lecture, the student may be largely engaged in reproducing the tutor's notes – a low-level skill. Likewise, if the year 1 question topic has not been 'taught',

the process of researching the answer may be at least as, if not more, demanding. It is for this reason that Knight (2006) asserts that we can only ensure validity at the local level where we are able to judge the quality of the learning processes behind the assessment products. This discussion alerts us to the potential limitations of what may appear to be demanding assessment tasks. A useful review of taxonomies of thinking skills can be found in Moseley *et al.* (2005), and Chapters 11 and 12 of this book discuss the relevance for designing assessment of the different 'levels' in the UK Framework for Qualifications in Higher Education (Quality Assurance Agency 2006d).

Constructive alignment

The notion of matching course objectives with assessment underpins the concept of 'constructive alignment' (Biggs 1996). It is a course design methodology which emphasises the centrality of intended learning outcomes. These should determine the teaching and assessment such that they are compatible and act to support one another. 'You wouldn't lecture students on how to teach using small groups, and give them a written test. You would get them to participate in small groups, then run their own and see how well they did it' (Biggs 2003: 27).

Constructive alignment, and the taxonomies discussed earlier, assume we can devise learning outcomes in a meaningful way (see Box 2.3. for an introduction to the debate). Although this is a contested proposition, this book has adopted an outcome-based approach to assessment design, on a pragmatic basis, because such an approach is widely employed.

Overall, the research suggests that tutors do need to be mindful that their assessment methods may not be as discriminating as they hope they are and may permit students to pass with 'conceptions of subject matter that teachers wished to change' (Ramsden 2003: 72) or to avoid large sections of the curriculum (Struyven *et al.* 2002). Chapter 13 on diversifying assessment provides practical ideas for increasing assessment validity, particularly in relation to examinations.

Authentic assessment

In addition, as discussed earlier, intrinsic interest in the learning activity is associated with deep approaches to learning. Students value assessment activities which appear worthwhile in themselves; they appear to have value beyond completing the task (Struyven *et al.* 2002). This notion has prompted the idea of 'authentic' assessment which is seen as 'more practical, realistic and challenging' (Torrance 1995). In reality, much of the move towards diversifying higher education assessment is based on the implicit notion of

Box 2.3 Leads into literature: the learning outcome debate

An outcome-based approach to specifying the curriculum has gained currency internationally in recent years. It aims to improve transparency by stating what achievements are represented by individual academic awards. It also provides for flexibility, for example by facilitating credit transfer and accreditation of prior learning (Gosling and Moon 2002).

Unfortunately, the term 'learning outcome' (LO) is often interchangeable with other educational constructs such as 'objectives', 'competencies', 'achievements' and 'skills', and this causes some confusion for educators (McGourty *et al.* 1999). In an attempt to clarify the term, McGourty *et al.* (1999: 4) define LOs as 'observable and measurable manifestations of applied knowledge', something which is 'reflected through the action and behaviour of the individual' rather than their ability to write about it. This is why some form of taxonomy of learning outcomes is useful; it lays out the types of cognitive skills students should be performing at different levels.

Hussey and Smith (2002) consider the specification of LOs to be part of a new regime of accountability. Outcome-based descriptions of academic awards notionally allow scrutiny of the appropriateness of the curriculum, how it reflects the subject benchmark statements and the standards being set. It allows comparison within and between institutions (Gosling and Moon 2002). This leads to one of the central criticisms of LOs, that they imply that the 'fuzzy' business of learning and teaching can be clearly specified (Hussey and Smith 2003). On the contrary, it is argued that they are difficult to write in a meaningful way (Knight and Yorke 2003) and have to be interpreted in relation to the context, which can only be done by those who are already familiar with that context (Hussey and Smith 2002). Thus they are not 'objective' or easily understandable by other stakeholders such as students.

Learning outcomes are also criticised for ignoring the unpredictable and indefinable aspects of learning, the 'emergent outcomes' (Hussey and Smith 2003: 362), and for reducing the value of subject content compared with skills. They may restrict learning as students focus on what is needed to pass. Jackson (2000) points to further criticisms, some more practical than educational, such as the time needed to prepare staff for the LO approach to course design and teaching, the bureaucracy, and the threat to professional autonomy contained in the requirement to make explicit what has been implicit.

However, there are strong arguments in favour of the outcome-based approach as opposed to former methods. Gosling and Moon contend that undefined learning is difficult to assess fairly, and students should know what is expected of them. Many academics associate such transparency with fairness, better communication and avoiding confusion between staff and students as the rules and structures are made clear (Orr 2005). Indeed Hussey and Smith (2003) do not reject the concept of LOs but urge that they are only written in general terms so that modules can embrace outcomes that emerge during the messy business of learning.

Gosling and Moon accept that we should avoid detailed lists of competences in higher education and use a broader description of outcomes, including skills, knowledge and understanding. In addition, Biggs (2003) suggests that the nature of higher-order outcomes means that they do not discourage unforeseen or unforeseeable outcomes because they tend to specify the process the student should demonstrate, rather than the detailed content.

Whatever your views, the learning outcome approach to documenting higher education is here to stay, at least for the foreseeable future. However, it may be wise to accept that learning outcomes can only be described in fairly general terms and must be interpreted in the light of the specific programme or module context.

authentic assessment, and higher education has a tradition of using it in various ways, for example on vocational awards. Nevertheless Knight and Yorke (2003) provide a convincing critique of authentic assessment both in terms of how it is being interpreted and in the pragmatic and economic costs of doing it well in the current regulatory framework of higher education. They do, however, recognise that the relative freedom of higher education gives more scope for authentic assessment than exists in schools.

Conclusion

This chapter has attempted to digest key themes in assessment research as a basis for the enhancement of practice. Essentially, they are as follows:

- Assessment strongly influences students' learning, including what they study, when they study, how much work they do and the approach they take to their learning.
- The type of assessment influences the quality and amount of learning achieved by students.
- Poorly designed assessment can lead to students developing limited conceptual understanding of the material, although the limitations of the assessment tool and process may mask this failure.
- Well-designed assessment is likely to be intrinsically motivating for students and lead to better retention of material which the students can apply in other settings.
- Students' prior experience of learning and perceptions of assessment may override attempts by lecturers to change their approach to learning, and they should be helped to a better understanding of assessment tasks.
- Assessment tasks may not be assessing what we think they are assessing, they may be assessing lower-level understanding of the

30 DEVELOPING EFFECTIVE ASSESSMENT IN HIGHER EDUCATION

material, and may be failing to assess the stated outcomes of a programme of study.

- Anxiety-provoking assessment is associated with a surface approach to learning by students.
- Feedback is the most important aspect of the assessment process for raising achievement, yet currently students express considerable dissatisfaction with much feedback and it does not always impact on their learning.
- Self- and peer assessment are crucial elements of helping students to learn from their assessment and become more autonomous learners.
- Feedback should inform tutors' teaching and support strategies as well as student activity.

3 The conflicting purposes of assessment

The purpose of this chapter is to discuss the various purposes of assessment. This may seem a rather academic debate, of little practical use to staff wishing to develop their module and programme assessment strategies. Nonetheless, once tutors attempt to seriously evaluate and enhance assessment methods, it becomes clear that there is a constant need to juggle these different intentions and the concentration on some purposes over others can distort the value of assessment in universities. Paying attention to neglected purposes helps pinpoint where our enhancement efforts should lie. Box 3.1 illustrates the sort of dilemma that the different purposes of assessment create.

The lecturer in case study 1 has designed an assessment which is assessment *as* learning (the students are doing the learning as they work on the assessment), assessment *for* learning (the peer and group learning encourages formative discussion and feedback) and assessment *of* learning in that the task attempts to be a valid measurement of the learning outcomes in terms of understanding and application of the law (see discussion of these terms in Chapter 2).

The moderator, on the other hand, emphasises her responsibility to focus entirely on the assessment *of* learning: whether this assignment is generating evidence of individual student achievement at the appropriate standard. Can she, with integrity, report to the institution that the marks awarded reflect the students' work and are comparable with standards on equivalent courses in other institutions? There is a tension between these different purposes of assessment which needs to be resolved.

First, let us consider four purposes of assessment:

1. **Certification:** providing the means to identify and discriminate between different levels of achievement, and between students, providing a licence to practise in the case of professional programmes, enabling selection of students for further study and employment – assessment *of* learning.
2. **Student learning:** promoting learning by motivating students, steering their approach to learning and giving the teacher useful information to inform changes in teaching strategies – assessment *for* and *as* learning.
3. **Quality assurance:** providing evidence for relevant stakeholders

32 DEVELOPING EFFECTIVE ASSESSMENT IN HIGHER EDUCATION

(for example employers, inspectors, external examiners) to enable them to judge the appropriateness of standards on the programme (Gibbs 1999) – assessment *of* learning.

4. **Lifelong learning capacity:** encouraging students to develop 'knowledge, skills and predispositions to underpin lifelong learning' (Boud 2000: 151) – assessment *as* learning.

Although each of these purposes appears entirely proper, they often conflict with one another. Boxes 3.2 and 3.3 illustrate this problem. While some assessment methods give confidence that the work is the student's own, other techniques promote higher-level learning. Some techniques provide reliable scores while others produce widely varying grades from different markers. In effect, the different purposes emphasise different principles of assessment.

The fourth purpose of 'lifelong learning capacity' does not always feature in lists of assessment purposes. However, it has been included here in recognition of the fact that assessment may also be important for developing

Box 3.1 Case study 1

A law tutor has introduced a simulation of an aspect of legal practice where student teams negotiate towards settlement of a dispute as if they were partners in a law firm. They spend three weeks working on these cases, posting their claims or defence against the claims on a group discussion page on the module's site within a virtual learning environment. The tutor e-moderates the discussion, responding with comments, questions and sources for further investigation. At the end of the period, each group presents the legal basis for their claim or defence in a classroom session. The student audience, working in threes, is asked to discuss the claim or defence and identify two strengths and two weaknesses of the legal argument adopted which are fed back to the presenting group. Each group assesses individual members' contribution to the team's work. The module tutor gives the presentation a mark which is altered for each member of the group depending on their peer assessment. The latter often, but not always, indicates that everyone should get the same mark.

However, the colleague moderating the module is not happy with the allocation of marks for two reasons. First, she feels that she has not been able to view the presentations and therefore cannot confirm that the marks given reflect the standard of work. Second, and most importantly, she thinks that it is impossible to guarantee that each student deserves the mark they gained. How does the tutor know that they didn't just turn up at the presentation and persuade the rest of the group to allocate them the same mark? She would prefer that the element of this exercise that counts towards the final mark is limited to individual students' written accounts of the legal negotiations.

Box 3.2 Case study 2

The summative assessment of a year 2 undergraduate philosophy module is a three-hour unseen examination, with a rubric requiring four answers, three from section A and one from Section B. Section A has a choice of eight questions and section B has a choice of three questions. The questions in section A require knowledge of the main topics on the course curriculum, whereas the questions in section B require students to draw learning from all topics of the course in order to analyse new material.

The students sit the examination in controlled conditions. Work is marked by the module tutor, with a sample moderated by a second tutor and marks adjusted accordingly. A sample of examination scripts is then sent to an external examiner who reports to the examination board on the comparability of marking standards.

In analysing the module results, the external moderator reports that standards are comparable, but he has asked why the students tend to achieve significantly lower marks in the section B essays compared with section A. Moreover, the tutor is concerned that section A tends to lead to regurgitation of his lecture material in a fairly undigested form.

Box 3.3 Case study 3

A year 1 undergraduate engineering module involves students in completing four on-line multiple choice question (MCQ) tests. The test questions have been developed and scrutinised for validity and accuracy by the course team. These tests must be taken during four one-week periods evenly distributed across the semester and each counts 10% towards the final module marks. The students sit the tests by making appointments to attend a controlled PC lab. The topics in the tests are also examined in the final unseen written examination. The students have access to paper and pens to use while doing the tests but they cannot remove any paper from the examination room.

The students get immediate feedback on their performance online at the end of the test, with the correct answer explained.

The tutor has seen an improvement in the end-of-course examination results, although he finds it difficult to write MCQs which test procedural rather than factual knowledge.

students' evaluative and self-assessment skills in preparation for employment and a future of learning (Jacques 2000; Boud and Falchikov 2006). This links to 'student learning' as a purpose of assessment, but stresses the active engagement of students in aspects of assessment, not just for the current

34 DEVELOPING EFFECTIVE ASSESSMENT IN HIGHER EDUCATION

course, but to help them develop the capacity to determine appropriate standards, discern the critical aspects of tasks, monitor their own progress, use feedback and other skills for lifelong learning (Boud 2000). Boud refers to this purpose as ‘sustainable’ assessment.

In order to analyse the conflicts illustrated by the case studies, it is necessary to examine them against a set of principles of assessment. There are varying views on what the principles of assessment should be (Jonsson and Baartman 2006; Quality Assurance Agency 2006c), and we have selected those which we see as particularly significant and which reflect the four different purposes listed earlier. Table 3.1 lists the different principles and evaluates each case study against them. It identifies how the assignments in the different case studies are robust in relation to certain principles and unconvincing in relation to others. The case studies also illustrate that achieving all of the principles is enormously difficult or impracticable. The remainder of this chapter discusses the implementation of each principle, followed at the end of the chapter (see Table 3.2) by suggested ways to resolve some of the weaknesses identified in the cases studies.

Validity

There is extensive debate in the educational literature about validity which we do not have the scope to review here. A range of different types of validity exist (Elton and Johnston 2002), many of which overlap with other principles such as effectiveness, reliability and transparency. For the purposes of this guide, we are focusing on ‘intrinsic validity’ (Brown *et al.* 1997), which means that **assessment tasks are assessing the stated learning outcomes** for the module, and this principle clearly underpins the notion of constructive alignment (see Chapter 2).

The traditional range of assessment tasks in higher education does not cope well with a wide range of learning outcomes – for example, those involving professional, subject-specific or **key skills** (Elton and Johnston 2002). Falchikov (2005) discusses the concept of ‘predictive validity’, the notion that an assessment tells us something about future behaviour. For example, on a nursing programme, does successful completion of an essay on the principles of care management predict whether the graduating nurse will have achieved the outcome of becoming a safe and effective practitioner in that sphere? In reality, we need an assessment process which more closely judges the students’ capacity to act as well as to express their factual and conceptual knowledge of the topic.

The discussion in Chapter 2 highlighted the problems associated with valid assessment in higher education, with difficulty increasing as the learning to be assessed moves towards higher levels of both dimensions of

Table 3.1 Analysis of case studies in assessment

<i>Principle</i>	<i>Link to purpose of assessment</i>	<i>Case study 1</i>	<i>Case study 2</i>	<i>Case study 3</i>
Validity	Student learning, certification	<p>Strong: The cases presented to the students require the ability to apply understanding of the law to legal problems. The exercise also assesses working as a team, solving complex problems and communicating effectively.</p>	<p>Weak: Results suggest students are focusing on the content of topics rather than understanding the underlying concepts that the tutor intended them to learn. May be testing examination technique.</p>	<p>Medium: Tests assess knowledge outcomes but are less successful at assessing higher-order outcomes concerned with application.</p>
Reliability	Certification, quality assurance	<p>Weak: One marker, without second marking or moderation. Marking open-ended problems has low levels of reliability.</p>	<p>Medium: Second marking and moderation take place to sector standards. Marking open-ended essay questions has low levels of reliability.</p>	<p>Strong: Testing instruments and marking independent of assessors.</p>
Effectiveness	Student learning, lifelong learning	<p>Strong: The task promotes learning through encouraging student activity and interaction, providing a motivational context by offering an authentic task and affording a means to help students structure their knowledge through working</p>	<p>Weak: Encourages students to memorise material rather than develop understanding. Summative examination limits opportunity to learn from feedback</p>	<p>Medium: Regular testing is likely to capture students' time. Immediate feedback is likely to support learning. The method may encourage rote memorisation of material.</p>

<i>Principle</i>	<i>Link to purpose of assessment</i>	<i>Case study 1</i>	<i>Case study 2</i>	<i>Case study 3</i>
Comparability and consistency	Quality assurance	<p>on an encompassing case study.</p> <p>Weak: The task generates little evidence for comparison across other courses. The open-ended nature means it is difficult to judge how the task should be weighted in comparison with other assignments or examinations</p>	<p>Strong: The course generates evidence of the quantity of assessment and the level of student achievement. It is possible to judge if assessment load is comparable with other cognate disciplines in the institution.</p>	<p>Strong: The course generates evidence of the quantity of assessment and the level of student achievement. It is possible to judge if assessment load is comparable with other cognate disciplines in the institution.</p>
Transparency	Student learning, quality assurance, lifelong learning	<p>Strong: The formative feedback from the tutor, participation in judging other students' work and giving and receiving feedback on contribution to the group all promote skills of evaluation and self-assessment.</p>	<p>Medium: Clear information is available on examination rubric and contribution to module mark. Students have access to previous papers but assessment criteria are implicit.</p>	<p>Strong: Clear information is available on test process, content and contribution to module mark. Repeat tests enable students to develop clarity regarding assessment method. Assessment criteria are explicit.</p>
Practicability	Student learning	<p>Strong: Student workload is managed by class time devoted to team meetings and use of a virtual learning environment to facilitate</p>	<p>Medium: Student workload is acceptable although the examination contributes to end-of-semester bunching of assessment. Staff marking</p>	<p>Strong: Assessment method distributes student workload across module. Test design is staff-intensive but automated marking</p>

Principle	Link to purpose of assessment	Case study 1	Case study 2	Case study 3
Equity	All purposes of assessment	<p>communication. Marking group presentations is swift</p> <p>Weak: Students with disability or English as an additional language may have more difficulty contributing to a group learning situation and/or communicating their learning in a presentation format.</p>	<p>workload concentrated although faster than coursework because detailed feedback is not required.</p> <p>Weak: Time-constrained examinations may pose difficulties for students with certain specific learning difficulties. Additional time may be given but this does not necessarily help (Robson 2005)</p>	<p>significantly reduces staff workload, of great benefit for large student cohorts.</p> <p>Medium: The test format can be altered to assist students with some disabilities. No time limit on test supports students with specific learning difficulties and those using an additional language.</p>
Attribution	Certification	<p>Weak: The tutor relies on the group members to peer-assess contributions to the team. Students could collude in misrepresenting individual contributions</p>	<p>Strong: Controlled conditions enable robust evidence that scripts are students' own work.</p>	<p>Strong: Controlled conditions enable robust evidence that tests represent student's own achievement. Procedure must ensure students can only log on in their own name.</p>
Conclusion		<p>Strong on student learning and lifelong learning. Weak on certification and quality assurance</p>	<p>Strong on quality assurance, medium on certification and student learning, weak on lifelong learning</p>	<p>Strong on quality assurance and certification, medium on student learning and lifelong learning</p>

38 DEVELOPING EFFECTIVE ASSESSMENT IN HIGHER EDUCATION

Anderson and Krathwohl's (2001) taxonomy (see Table 2.1). Case study 2 (Box 3.2) about the philosophy examination is a good illustration. Students may successfully compose answers to examinations because they have learned the relevant knowledge as organised for them by the tutor, rather than because they have grasped the meaning of key concepts in such a way that they can apply them to new situations. In general, examinations are valid for assessing recall and understanding of factual and conceptual knowledge.

Overall, the biggest threat to validity arises from attempts to ensure 'reliability' and 'practicability' (see the relevant sections below).

Reliability

Assessment tasks should be generating comparable marks across time, across markers and across methods. For example, reliability is demonstrated when different markers make the same judgements about an assignment or when one marker makes consistent judgements about a piece of work at different times. Overall, despite the QAA (2006c) urging higher education institutions to implement principles and procedures for reliable assessment, the evidence on this matter is depressing (Elton and Johnston 2002; Falchikov 2005), with little evidence of reliability in the marking of written work in higher education. Murphy (2006) argues that reliability (dependability) of assessment in UK higher education has traditionally been of little concern, with assessment decisions largely confined to those who have taught the students. Knight (2001) describes reliability in higher education as 'ramshackle' in comparison with secondary/high school education.

In addition, reliability requires that assessment of the same learning by different modes should render similar outcomes. Research on the latter hardly exists, although there have been studies comparing student achievement on two examination papers (Elton and Johnston 2002) which suggested the papers did not produce reliable measures of the same learning outcomes. Yorke *et al.* (2000) show a similar discrepancy in relation to the difference between examination and coursework and argue that the two methods must be assessing different things if the results are so consistently different.

The causes of unreliability are related to the nature of what is being measured by assessment in higher education. Knight (2006) argues that complex learning cannot be reduced to something simple enough to measure reliably; the more complex the learning, the more we draw on 'connoisseurship' (Eisner 1985) rather than measurement to make our judgements. However, he does accept that achievements in some disciplines may be more determinate, and therefore more open to reliable judgement. On the other hand, Maclellan (2004a: 312) would reject many assignments, for

example essays, as contradicting the fundamental belief 'that there can be universality of meaning as to what any grade or score represents'. An interpretivist view (see Box 3.4) would argue that there is a level of professional judgement in some elements of undergraduate assessment whatever the discipline, particularly if we take Knight's (2006) view that all graduates should be learning complex capabilities such as initiative, adaptability and critical thinking. According to Broad (2000), conferring grades in complex written work is impossible and misleading. It is interesting to note that the Quality Assurance Agency in the UK, in its recent code of practice for assessment (2006c), is advocating broader grades or mark bands, perhaps recognising the difficulty of precise percentage grading.

Box 3.4 Leads into literature: assessment policy and practice

Elton and Johnson (2002) provide an excellent discussion and review of the literature in relation to the key dilemma in higher education between assessment for certification and assessment for learning, setting out the different positions of the positivist and the interpretivist approaches. Essentially, a positivist approach believes in the importance of validity and reliability, assuming that objective standards can be set. The alternative, interpretivist, approach rejects objective truth and conceives of assessment as based on a local context, carried out through the judgement of experts in the field. In their view, it is a social practice whose credibility emerges from a community of practice which shares a consensus about what constitutes accepted knowledge, rules and procedures. It is a 'good enough' (2002: 39) approach in which 'dependability is parallel to reliability in positivist assessment' (2002:46)

On a more general front, Knight and Yorke (2003) provide a well-argued challenge to much current assessment practice, and Falchikov (2005) presents a useful critique of traditional assessment practices and the paradigm of assessment as quality control. Heywood (2000) offers a wide-ranging review of the literature on principles and practice in higher education assessment.

Academics, as 'connoisseurs', are considered able to make expert and reliable judgements because of their education and socialisation into the standards of the discipline and of their local context (Ecclestone 2001). Knight (2006) argues that this situated and socially constructed nature of standards means that measurement of complex learning can only take place effectively within its context, a local judgement made within that social environment, be it a teaching team, department or subject discipline. This local nature of standards means it is unsurprising that many studies have found considerable marking discrepancies between tutors. This is discussed in depth in Chapter 6.

While reliability is particularly important for the purposes of ‘certification’ and ‘quality assurance’, it may well work against validity. For example, an overriding concern for demonstrably reliable marking may prevent the use of group assignments or may encourage use of assessments that usually foster low-level learning such as multiple choice question tests (Scouller and Prosser 1994). However, quite reasonably, positivist approaches to assessment would argue that without reliability, there is no validity (Elton and Johnston 2002).

Overall, there has to be a ‘trade-off’ between reliability and validity, and, whereas American universities have erred towards reliability, validity has taken precedence in the UK (Hornby 2003). Perhaps the best we can do is attempt to achieve a reasonable balance across a student’s programme. The concern with moderation and reliability as a key element of quality assurance and certification is discussed further in Chapter 8.

Effectiveness

Assessment tasks should be designed to encourage good quality, ‘deep’ approaches to learning in the students. Various researchers list different purposes for assessment (Gibbs 1999; Elton and Johnston 2002; Dunn *et al.* 2004), several of which are directly related to assessment as an effective tool for promoting learning. They include ‘capturing student time and attention, generating appropriate student learning activity, providing timely feedback which students pay attention to, helping students to internalise the discipline’s standards and notions of quality’ (Gibbs 1999: 47) as well as motivation and preparation for life (Elton and Johnston 2002). Dunn *et al.* (2004) add the purpose of diagnosing student difficulties. Chapter 2 has set out the research on the relationship between assessment and student approaches to learning and achievement, and there is considerable consistency within the literature on what is required for assessment design to support effective learning.

Overall, Elton and Johnston (2002: 39) link effectiveness to validity: ‘Newer notions of validity stress that a “valid” procedure for assessment must have a positive impact on and consequences for the teaching and learning’. Black and Wiliam (1998a) argue that in general the grading function of assessment is overemphasised in comparison with the learning function.

Comparability and consistency

There should be consistent and comparable approaches to the summative assessment requirements of awards of the same level across programmes and institutions. Interpretations of this differ, but

normally it involves consistency to be shown in the level of learning demonstrated and in the learning hours required to complete the assessment for a particular number of **credits**. 'The primary concern should be whether standards compare favourably with those in similar institutions or against appropriate benchmarks' (Dunn *et al.* 2004: 69). Within programmes, consistency implies that modules offering similar credit must have comparable assessment requirements. This need not, necessarily be interpreted as exactly similar word lengths and should not mean similar types of assessment. It is the level of learning to be demonstrated and the notional hours required to complete the assessed tasks that are important.

Comparability and consistency also apply to multiple campus and collaborative provision, ensuring that students studying for the same awards have a comparable experience with similar outcomes, workload and standards applying across the different sites. Another element contributing to this principle is consistency in determining how individual assignment and module marks combine to generate an overall classification of award such as an upper second class degree, merit or grade point average, but there is clear evidence in the UK that practices vary significantly between and within universities (Yorke *et al.* 2004; QAA 2006a), leaving degree classification unreliable and inconsistent (Elton, 2004). Above all, comparability and consistency are related to ensuring quality standards and fairness rather than a direct link to assessment for learning.

Comparability and consistency apply specifically to summative assessment and are regulated through course approval, second marking and external monitoring, with all the costs and bureaucracy involved. As Gibbs (2006b) points out, the difficulty of engaging those students with a strategic approach to their learning has led staff to make all assignments summative, which means the full panoply of expensive procedures comes into force with precious resources diverted away from teaching.

Equity

Students enjoy equal opportunity to effectively demonstrate their learning. There is an overriding need to improve the performance of higher education in promoting equity, and assessment has its part to play in that drive. This principle may be considered from a compliance point of view, for example ensuring that practice complies with legislation. Alternatively, programmes and institutions can take a proactive approach which encourages students to celebrate and express diversity and difference within their assessed work.

Chapter 10 on supporting widening participation and student diversity tackles this principle in a practical way. It recommends that, in general, good assessment practice is also inclusive assessment practice. Implementing

42 DEVELOPING EFFECTIVE ASSESSMENT IN HIGHER EDUCATION

equity includes making reasonable adjustments for students' special educational needs and/or disabilities and ensuring cross-institutional consistency in assessment procedures such as **extenuating circumstances** claims, providing for specific learning needs (see precept 5 in QAA 2006c), blind **second marking** and anonymous marking.

The principle of equity is important for all the four purposes of assessment – certification, student learning, quality assurance and lifelong learning capacity.

Practicability

Assessment tasks must be practicable for both staff and students in terms of the time needed for completion and marking. There is always likely to be a trade-off between other principles and practicability. Knight (2006), for example, in discussing students' performance as professionals, points out that reliable judgements can only be made when there have been several observations by multiple observers in a range of contexts, which is not very practicable in terms of resources. Practicability also relates to the amount and timing of assessment, and the need to avoid bunching of assessment deadlines. Availability of staff, venue, equipment and technical support for examinations and assignments is also an important issue of practicability. Imaginative and valid assessments have come unstuck at the point of finding sufficient space, staff or equipment to operate.

Modular degrees have brought concerns about over-assessment. Ross (2005) argues that streamlining assessment is necessary to enable staff and students to cope with the changing environment in higher education, particularly growth in student numbers. He also proposes 'stripping-back' of assessment to 'its fundamental relationship with learning outcomes' as a means of making it practicable in the contemporary context. Chapters 12 and 13 discuss methods of managing the student assessment workload.

Transparency

Information, guidance, rules and regulations on assessment should be clear, accurate, consistent and accessible to all staff, students, practice teachers and external examiners. This principle has been of growing importance over the last decade in the UK as institutional processes have become more clearly subject to external audit and review. Transparency is seen as a key element of institutional accountability (Orr 2005), where clear procedures and rules facilitate external scrutiny. These include identifying learning outcomes and marking criteria, extension and extenuating

circumstances rules, and second marking and moderation procedures. While much of this may not necessarily improve learning from assessment, it does allow the system to be 'judged in relation to its overall coherence and transparency' (Crook *et al.* 2006: 96), suggesting that reliability in assessment has been substituted by a concern for reliability of assessment procedures (Brown and Knight 1994).

On a more positive note, many academics associate transparency with fairness, better communication and avoiding confusion between staff and students as the rules and structures are made less opaque (Gosling and Moon 2002; Orr 2005). Sadler (2005) supports the drive towards transparency, noting that mysterious criteria and standards lead to a relationship of dependency between student and tutor. Students rely on tutor judgement because they have no concept of how their work is being judged, leaving them unable to regulate their own learning.

The drive towards transparency is seen in various forms. UK institutions are required to write and publish programme specifications. In relation to assessment, it is the view of the QAA (2006c) that clear, public marking criteria and **marking schemes** are important in ensuring that marking is carried out fairly and consistently across all disciplines. However, evidence discussed under 'reliability' would suggest that transparency in relation to 'complex' assignments is an enormous challenge. Orr (2005: 178) argues that tutors develop ideas and beliefs about assessing students which determine how they judge work and 'this implicit approach may contradict the explicit requirements' given to students. Price (2005) cites a range of studies where variation between staff continued despite the use of assessment criteria.

Furthermore, information such as assessment criteria provides limited help to students as they require interpretation within the context of the given situation (Crook *et al.* 2006). As a consequence, O'Donovan *et al.* (2004) argue that, rather than dispense with 'transparent' information, teachers in higher education need to use participative methods in order to help students learn the tacit knowledge associated with assessment, to begin to make the hidden more visible (see Chapter 5).

A further feature of transparency is demonstration of fair and transparent mechanisms for marking and moderating marks (QAA 2006c) so that all stakeholders are able to judge their appropriateness. Case study 1 is an example of a response to this pressure. The moderator's concern for evidence that the students have individually met the learning outcomes entails time spent by students (writing) and staff (marking) which may add little in terms of student learning.

Overall, transparency is important for 'student learning' but also for 'quality assurance' because of the links to institutional accountability. In the UK, quality assurance is based on the academic infrastructure which is controlled by the QAA (2006d).

Attribution

Tasks should generate clear evidence that the work (of whatever nature) has been produced by the candidate. This relates to a range of malpractice including plagiarism but also collusion – for example, where a student team is prepared to represent one or more of their members' contributions inaccurately. This principle has come to the fore in recent years as the internet and electronic communication have significantly increased concerns about plagiarism and cheating.

This is important for 'certification' as institutions and external stakeholders, quite reasonably, want to be assured that students have achieved the learning for the award they have gained. Stringent efforts are gradually developing across the higher education sector to tackle malpractice, with a holistic approach recommended which includes both prevention and detection (MacDonald and Carroll 2006). The emphasis in this book will be on incorporating malpractice prevention into the teaching of modules (Chapter 4) and the design of assessment strategies which help to prevent it (Chapters 11 and 12).

Conclusion and possible ways forward

The foregoing discussion has illustrated the conflicting nature of established principles underlying assessment practice. If we link these principles back to our initial discussion of the purposes of assessment, we can see that each purpose emphasises different principles and is hindered by others:

1. **Certification** emphasises validity, reliability, equity and attribution.
2. **Student learning** emphasises validity, effectiveness, practicability, equity and some aspects of transparency.
3. **Quality assurance** needs validity, reliability, transparency, equity, comparability and consistency.
4. **Lifelong learning** emphasises validity, effectiveness, equity and elements of transparency.

Therefore, in designing our assessment strategies, we need to be aware of the different purposes and principles of assessment and the need to achieve balance. Traditionally assessment has focused on measurement, the summative assessment *of* learning (Boud 2000), and most institutions still stress the measurement aspects of assessment with little concern for other purposes (Gibbs and Simpson 2004–5; Hounsell 2006). In addition, in the current era of

accountability, universities are now placing considerable emphasis on the quality assurance aspects of assessment with comparable, consistent and transparent procedures which may also have little to do with supporting learning.

However, researchers are now stressing the importance of balancing concerns about assessment *of* learning (certification and quality assurance) with assessment *for* and *as* learning (student learning, lifelong learning), as discussed in Chapter 2. 'We are not arguing for unreliable assessment but we are arguing that we should design assessment, first, to support worthwhile learning, and worry about reliability later' (Hounsell *et al.* 2006: 1). Of course, this idea of abandoning the quest for reliable and transparent university assessment in an age of accountability would be institutional suicide, and it does not take great powers of imagination to predict the newspaper headlines. Consequently, a pragmatic approach has to be one of balance across a student's programme. It has to be one where care is taken in assignment design to reconcile the potential contradictory effects of different assessment purposes and underlying principles. This is a key proposition underpinning the content of this book as the different principles discussed earlier are developed in relation to all aspects of the assessment cycle.

That process is commenced here by returning to our case studies in order to consider ways to tackle the shortcomings that they illustrate (see Tables 3.2–3.4).

Table 3.2 Case study 1: group presentation

<i>Principle</i>	<i>Potential enhancements</i>
<i>Reliability</i>	Use two markers for the student presentations. Markers (and students) agree marking criteria in advance and share understanding by marking a formative presentation or video of a presentation.
<i>Comparability and consistency</i>	Staff estimate learning hours required by students to complete the assessment and compare these with learning hours for alternative assessments in order to judge weighting.
<i>Equity</i>	Consider putting international students in common language groups to ease communication. Discuss requirements with disabled students and identify any reasonable adjustments.
<i>Attribution</i>	Students are required to keep 'minutes' of their group meetings identifying attendance, decisions made and action taken. These allow the tutor to gain some indication of individual contributions. Alternatively, group pages are set up on virtual learning environment and the tutor can check individual contributions to the discussion.

Table 3.3 Case study 2: unseen examination

<i>Principle</i>	<i>Potential enhancements</i>
<i>Validity</i>	Change examination paper to part B type questions only. Give no choice but provide exemplar questions early in the programme so that students direct their efforts towards the intended learning.
<i>Reliability</i>	Develop assessment criteria and marking scheme in discussion with module tutors. Staff team pre-mark sample scripts to discuss and agree interpretation of marking scheme.
<i>Effectiveness</i>	Change examination as above but include mock examination part-way using self- and peer assessment to help students explore their achievement against standards.
<i>Transparency</i>	Publish assessment criteria and marking scheme. Peer and self-assessment of mock paper to help students better understand the scheme.
<i>Practicality</i>	Consider shorter examination with focus on key concepts to assess whole course but reduce marking.
<i>Equity</i>	Consider second mode of assessment for module which offers students an alternative way to demonstrate their learning.

Table 3.4 Case study 3: online multiple choice quiz

<i>Principle</i>	<i>Potential enhancements</i>
<i>Validity</i>	Ensure that the second mode of assessment for the module focuses on application of knowledge, perhaps through an investigation or project.
<i>Effectiveness</i>	Generate each test uniquely from question banks and allow students to sit a test as many times as they like in each week with only the last attempt counting – to encourage students to engage with feedback. Use second assessment as above to encourage deep approach to subject matter.
<i>Equity</i>	Discuss test format with disabled students and disability officer or learning technologist to identify appropriate adjustments.
<i>Attribution</i>	Procedure must ensure students can only log on in their own name.