Improving Assessment Through Student Involvement

Practical solutions for aiding learning in higher and further education

Nancy Falchikov
Improving Assessment Through Student Involvement

The assessment of students is an activity central to the role of any professional in further and higher education, and is an area that is the subject of constant innovation and debate.

This book provides a scholarly account of the many facets of assessment, with a particular focus on student involvement. Peer and self-assessment are powerful assessment tools to add to the existing tutor-based methods of assessment and feedback, and this book is a comprehensive guide to the methods and issues involved.

Practical and accessible in style, yet grounded in research and rich in evidence-based material, Improving Assessment Through Student Involvement will be valued by all FE or HE professionals wanting to enhance both the effectiveness and quality of their assessment methods.

Nancy Falchikov is a respected authority on peer assessment, and worked for many years as a teacher and researcher in higher education. She has written widely on assessment matters, and is the author of Learning Together: Peer Tutoring in Higher Education (RoutledgeFalmer).
Improving Assessment Through Student Involvement

Practical solutions for aiding learning in higher and further education

Nancy Falchikov
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>viii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>xii</td>
</tr>
<tr>
<td><strong>1 The seven pillars of assessment</strong></td>
<td>1</td>
</tr>
<tr>
<td>Why assess?</td>
<td>1</td>
</tr>
<tr>
<td>How to assess?</td>
<td>4</td>
</tr>
<tr>
<td>What to assess?</td>
<td>8</td>
</tr>
<tr>
<td>When to assess?</td>
<td>22</td>
</tr>
<tr>
<td>Who does the assessing?</td>
<td>27</td>
</tr>
<tr>
<td>How well do we assess?</td>
<td>28</td>
</tr>
<tr>
<td>Whither? What do we do or where do we go next?</td>
<td>31</td>
</tr>
<tr>
<td><strong>2 What’s wrong with traditional assessment?</strong></td>
<td>32</td>
</tr>
<tr>
<td>Limitations of assessment as measurement</td>
<td>32</td>
</tr>
<tr>
<td>Reliability and bias in teacher and examiner marking</td>
<td>34</td>
</tr>
<tr>
<td>Negative side effects of traditional assessment</td>
<td>36</td>
</tr>
<tr>
<td>The relationship between traditional assessment and academic dishonesty</td>
<td>40</td>
</tr>
<tr>
<td>The role of the Internet in facilitating and detecting cheating</td>
<td>52</td>
</tr>
<tr>
<td>Advice to practitioners: strategies for preventing cheating</td>
<td>54</td>
</tr>
<tr>
<td>Summary</td>
<td>57</td>
</tr>
<tr>
<td>Appendix: Internet resources to help combat plagiarism</td>
<td>58</td>
</tr>
<tr>
<td><strong>3 Changing definitions of assessment</strong></td>
<td>59</td>
</tr>
<tr>
<td>Assessment as measurement</td>
<td>60</td>
</tr>
<tr>
<td>Assessment as procedure</td>
<td>60</td>
</tr>
<tr>
<td>Assessment as enquiry</td>
<td>62</td>
</tr>
<tr>
<td>Assessment and accountability</td>
<td>62</td>
</tr>
<tr>
<td>Assessment as quality control</td>
<td>63</td>
</tr>
<tr>
<td>Chapter</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------------------------------------------</td>
</tr>
<tr>
<td>4</td>
<td>Why do teachers involve students in assessment?</td>
</tr>
<tr>
<td></td>
<td>Themes from the 1950s to 1980s</td>
</tr>
<tr>
<td></td>
<td>The 1990s: benefits and pressures</td>
</tr>
<tr>
<td></td>
<td>The early 2000s: the rise of the machines?</td>
</tr>
<tr>
<td></td>
<td>Overview</td>
</tr>
<tr>
<td></td>
<td>Summary: benefits of involving students in assessment</td>
</tr>
<tr>
<td>5</td>
<td>How may students be involved in assessment?</td>
</tr>
<tr>
<td></td>
<td>Peer assessment</td>
</tr>
<tr>
<td></td>
<td>Self-assessment</td>
</tr>
<tr>
<td></td>
<td>Collaborative assessment</td>
</tr>
<tr>
<td></td>
<td>What is the level of student involvement?</td>
</tr>
<tr>
<td></td>
<td>What do students assess?</td>
</tr>
<tr>
<td></td>
<td>How are assessments carried out?</td>
</tr>
<tr>
<td></td>
<td>Recommendations to help combat bias in alternative assessments</td>
</tr>
<tr>
<td></td>
<td>Conclusion</td>
</tr>
<tr>
<td>6</td>
<td>Practical peer assessment and feedback: problems and solutions</td>
</tr>
<tr>
<td></td>
<td>Frequently asked questions</td>
</tr>
<tr>
<td></td>
<td>Peer feedback marking</td>
</tr>
<tr>
<td></td>
<td>Final comment</td>
</tr>
<tr>
<td>7</td>
<td>How well are students able to judge their own work?</td>
</tr>
<tr>
<td></td>
<td>A qualitative review of self-assessment studies</td>
</tr>
<tr>
<td></td>
<td>A meta-analysis of student self-assessment studies</td>
</tr>
<tr>
<td></td>
<td>Recent self-assessment studies</td>
</tr>
<tr>
<td></td>
<td>Conclusion</td>
</tr>
<tr>
<td>8</td>
<td>How reliable or valid are student peer assessments?</td>
</tr>
<tr>
<td></td>
<td>A meta-analysis of peer assessment studies</td>
</tr>
<tr>
<td></td>
<td>Results</td>
</tr>
</tbody>
</table>
Preface

What this book is about

This book is about how students have been, are, and may be involved in assessment. As the title indicates, it is my belief that the process of assessment itself may be improved by student involvement, and that student learning may also benefit. In Chapter 1, we look at the seven pillars of assessment, on which the canopy of assessment rests. These pillars apply to all types of assessment, not merely those involving students. Seven questions are answered, relating to why how and what we assess, when we carry out assessments, who does the assessing and how well it is done. The final question, ‘Whither?’, asks what do we might do and where we might go next.

Chapter 2 poses the question, ‘What’s wrong with traditional assessment?’ Some limitations of traditional assessment are considered, beginning with a review of reliability and bias in teacher and examiner marking. Not only is this issue important in its own right, it also has implications for testing the reliability or validity of self- or peer assessment. Typically, when students are involved in assessment, teacher marks are used as the standard against which student-generated marks are compared. Negative side effects of traditional assessment are discussed, including a brief exploration of the relationship between traditional assessment and academic dishonesty. The chapter concludes with a consideration of the role of the Internet in both facilitating and helping detect cheating, and advice to practitioners on this topic.

In Chapter 3, ‘Changing definitions of assessment’, conceptions are traced through three phases defined by Pearson et al. (2001); ‘assessment as measurement’, ‘assessment as procedure’ and ‘assessment as enquiry’. In addition, I argue that we have already entered a fourth phase, ‘assessment as quality control’, which seems to be co-existing with other phases, particularly assessment as enquiry. I ask how it is that assessment as enquiry and assessment as quality control co-exist. What is the balance between them? How may we identify and preserve the best of both systems?

The question of why teachers involve students is discussed in some detail in Chapter 4, ‘Why do teachers involve students in assessment?’. Reasons
given by teachers who have published their own work are inspected, beginning with the 1950s. The chapter ends with a discussion of the benefits of involving students.

Chapter 5, ‘How may students be involved in assessment?’, is central to the theme of this book. In it, I review the ways in which students may be involved. Considerable space is allocated to peer and self-assessment, but collaborative assessment and feedback are also discussed. The chapter then goes on to look at the level of student involvement and closes with advice for practitioners.

In Chapter 6, ‘Practical peer assessment and feedback: problems and solutions’, possible answers to a number of frequently asked questions are presented. In this chapter, I also describe the development of peer assessment called ‘Peer Feedback Marking’ which, as the name implies, emphasizes feedback rather than marks.

Chapter 7, ‘How well are students able to judge their own work?’ and Chapter 8, ‘How reliable or valid are student peer assessments?’ describe the conduct and results of two meta-analytic studies: one dealing with self- and the other with peer assessment. The meta-analytic technique is described briefly, and outcomes of the two studies discussed. Each chapter ends with recommendations for conducting studies that wish to maximize the agreement between student and teacher marks.

Assessment in and by groups is the topic considered in Chapter 9, ‘Assessment of groups by peers’. A number of studies are reviewed and the problems of peer assessment in groups discussed. Strategies for minimizing or avoiding problems are discussed and, once again, the chapter ends with recommendations for practitioners.

In Chapter 10, ‘Computer Assisted Assessment (CAA) and student involvement’, I survey some recent developments in Computer Assisted Assessment (CAA), identifying a number of functions. Many of these functions involve students – some actively, some less so. Software now exists which acts to support the assessment process itself, for example, by facilitating the collection and management of marks. Web-based testing is also taking place. More interestingly, several examples of computer-based self- and peer assessment have been identified and these are described, along with other examples of computer support for, and encouragement of, student learning. Again, advice to practitioners brings the chapter to a close.

Chapter 11 summarizes the key findings and identifies areas for further research.

**Who is this book for?**

This book is for anyone in further or higher education who is concerned to improve the practice of assessment generally or to provide their students with as rich an educational experience as possible. As we have
seen, it contains chapters on key issues, a critique of traditional practices and a survey of reasons why teachers choose to involve their students in assessment. It also provides a research-based overview of current practice of student involvement in assessment (self-, peer and collaborative assessment, assessment in groups and Computer Aided Assessment), and is, therefore, of interest to researchers in education as well as to practising
How to read this book

Your particular needs are likely to determine how, exactly, you will read this book. It may be that you will choose to dip in and out as your needs dictate. There are, however, two main routes through it. All readers will probably wish to orient themselves by looking at the first two chapters. Then, those concerned with issues relating to assessment generally will tend to follow what I have called the ‘Quality and standards route’, while those interested in specific ways of involving students will follow the ‘Involving students and improving learning route’.

The ‘Quality’ route contains Chapters 3, 7, 8 and 11. This route begins with a critique of traditional assessment and goes on to examine reliability and validity of self- and peer assessment.

The ‘Involving students’ route takes readers through Chapters 4, 5, 6, 9 and 10 before concluding with Chapter 11. After working through Chapters 4 and 5, readers following this route may wish to choose a particular path to follow. Those interested in peer assessment and its problems will wish to read Chapter 6, ‘Problems of peer assessment and peer feedback marking (PFM)’. This chapter is of particular relevance to teachers and staff developers. Those who wish to learn more about peer assessment in groups and to deal with problems of how to differentiate between individual contributions when awarding marks, will choose Chapter 9, ‘Peer assessment in groups’. Chapter 10, ‘Computer Assisted Assessment (CAA) and student involvement’ provides a third path. All these chapters contain practical advice for practitioners.

The structure of the book is summarized in Figure 0.1.
Acknowledgements

I wish to thank a number of people for help and support during the researching and writing of this book.

Thanks are due to the American Educational Research Association for their permission to reproduce the figures from *Review of Educational Research* (Volume 70, number 3, pages 305–6) in Chapter 8, and to Phil Davies for his screenshot in Chapter 10. Thank you, too, to Joanna Bull for permission to use extracts from the CAA Centre website materials, also in Chapter 10.

I am indebted to Mike Prosser at the University of Sydney for providing me with research facilities, and to other colleagues there for their welcome and stimulating discussions. Thanks are also due to David Boud and Judy Goldfinch with whom I worked on the two meta-analytic studies discussed. Thank you, Phoebe Fisher, for your illustration which we are, unfortunately, not able to include in the book. Thanks are also due to many other colleagues who have supplied me with copies of their own work or pointed out useful references: Dai Hounsell, Mark Lejk, Mary McCulloch, Fred Percival to name but four. Thank you Kim Addison for your practical support. George Brown once again provided encouragement and critical advice, as did Greg Michaelson, who has completed a double marathon of reading and critiquing. Thank you both. Finally, thank you Steve Jones for being a cat person.
Chapter 1

The seven pillars of assessment

This chapter considers the seven pillars on which the canopy of assessment in higher education rests. Careful attention to each of these pillars is necessary. The seven are

(1) Why assess?
(2) How to assess?
(3) What to assess?
(4) When to assess?
(5) Who assesses?
(6) How well do we assess?
(7) Whither? What next?

We shall now look at each.

1 Why assess?

This question deals not only with the purposes of assessment, but also the issue of who needs or uses the results. In 1987, Rowntree observed that the vast bulk of assessment literature is concerned with how to use assessment for purposes of grading and ranking. ‘Only a minuscule proportion considers how to use it [assessment] to enhance the students’ educational growth’ (Rowntree, 1987: 10). In exploring the purposes of assessment, Rowntree identified six broad categories:

(i) Selection
(ii) Maintaining standards – or quality control
The seven pillars of assessment

(iii) Motivation of students
(iv) Feedback to students
(v) Feedback to teachers
(vi) Preparation for life

To what extent and in what ways do these purposes support student learning? It can be argued that selection and quality control benefit stakeholders other than students, though students need to be assured of the quality of the awards they achieve. Rowntree (1987: 22) talks of the ‘constant prod from assessment’ which encourages learning. Thus, the motivational purpose may be said to be more directly related to the needs of students than other purposes of assessment. However, Rowntree argued that motivational assessment may be seen as an instrument of coercion, a way of getting students to do something they wouldn’t do otherwise. In this way, motivational assessment may also benefit the teacher rather than students. Thus, motivation has two faces. It encompasses both encouragement and coercion, the carrot and the big stick.

Feedback is more obviously beneficial to students, and perceived as such by them. Preparation for life, on the other hand, clearly depends on the sort of life you wish to lead. Traditional interpretations of ‘preparation’ have focussed narrowly on employment and career advancement and have been predicated on competition. However, there are those, myself included, who see collaboration and sharing as more valuable life skills than excellence in struggle and competition, and assessment’s role in preparing students for life seems to be changing. Birenbaum (1996), for example, stressed the function assessment has to help develop a self-regulated learner, capable of communicating and co-operating with others.

These categories form a useful framework within which to consider assessment, but, as we have seen, emphases are changing. While Rowntree observed that little assessment literature was concerned with enhancing student growth, this small proportion is increasing rapidly, and, as we shall see in Chapter 4, many more teachers now wish to use assessment to achieve benefits for their students.

**purposes of assessment in the context of group work**

Group working provides us with some additional purposes of assessment. This practice has become more widely used in recent years, and its assessment has often given rise to problems, most notably the difficulty of assigning individual grades for group efforts (e.g. Magin, 2001b). Although we shall return to discuss group work in the context of peer assessment in Chapter 9, we shall now consider how the purposes of assessment of group working may differ from those of learning alone. Webb (1995) considered theoretical and practical issues that need to be
taken into account in the design, use and interpretation of assessments carried out in collaborative small groups. In her discussion of some of the reasons given to justify the use of group work, she noted that the most often cited reasons did not always coincide with the purposes of assessment. She identified four such purposes in the context of group work:

1. to measure individual achievement (traditional purpose);
2. to measure performance after learning in group setting;
3. to measure group productivity and effectiveness;
4. to measure students’ abilities to work collaboratively.

Webb observed that these purposes sometimes represent competing goals of group work. For example, group productivity often competes with learning from group work. She also argued that the group processes may be different in the two contexts. ‘Behavior that is conducive to producing a high-quality group product may not always be conducive to individual learning, and vice versa’ (Webb, 1995: 241).

**Formative and summative assessment**

In answering the question, ‘Why assess?’, Biggs (1999) pointed to the important distinction between formative and summative functions. George and Cowan (1999: 1) described summative evaluation as consisting of judgements which, ‘for the immediate future, form the basis of one or more decisions which stem from that judgement’. In summative assessment, students are graded at the end of a module, unit or course. Summative judgements are also used to accredit learners at the end of a programme. Nevo (1995) characterized these judgements as being for diagnostic and certification purposes. George and Cowan (1999: 1) also saw evaluation as formative ‘when the intention is to identify scope and potential for improvement’. In formative assessment, results of assessment are used for feedback to teachers and students alike. Wiliam and Black (1996) identified feedback as a key component to formative assessment. They stressed the importance of Ramaprasad’s (1983) conceptualization of feedback as ‘information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way’. Thus, the authors reasoned that formative functions of assessment are validated in terms of their consequences. Hounsell (2003) provided an illuminating account of the tensions between our knowledge of the impact of formative feedback on learning and concerns about reduced opportunities for students to benefit from it.

Black and Wiliam (1998: 143) argued that formative assessment is at the heart of effective teaching and that self-assessment is ‘an essential component of formative assessment’ (italics in original). Their review of research
literature found evidence to support the assertion that formative assessment improves standards. In addition, many studies found that formative assessment helps low achievers more than other students.

Wiliam and Black (1996) characterized the relationship between formative and summative functions as a continuum. At the formative pole, assessments are evaluated by the extent to which they provide the basis for successful action and ignore problems of creating shared meanings. At the other, summative pole, shared meanings are important and ‘undesirable consequences that arise are often judged by appeal to the need to create consistency of interpretation’ (Wiliam and Black, 1996: 544). They concluded that significant tensions are created when the same assessments are required to serve both formative and summative functions. However, George and Cowan (1999: 17) saw the distinction between formative and summative evaluation as, ‘a balance, which in the early stages of development will emphasise formative rather than summative evaluation; and which, in later stages, will dwell on the reverse balance’.

**Who uses the results of assessment?**

Dietel et al. (1991) identified four main groups of users of assessment: policymakers, administrators, teachers and students (and, during compulsory education, their parents). Each group uses assessment for a different purpose. The key purposes of assessment and areas of overlap between different stakeholders are summarized in Figure 1.1.

Answers to the question, ‘Why assess?’ may be grouped into two categories: summative and formative. In the former group are purposes such as selection and certification, as well as accountability and effectiveness monitoring. Purposes in the latter group are more student (and teacher) centred, including diagnosis, motivation, feedback and improving learning.

**2 How to assess?**

There are a number of conflicting modes of assessment. They vary from the informal to the formal. Assessment may be carried out quantitatively or qualitatively. It may take place at the end of sessions or modules or be spread throughout the academic year. It may focus on the product or process of learning. We shall look at a number of contrasting modes.

**Quantitative vs qualitative**

Biggs (1999) elaborated this distinction (see Table 1.1).

Quantitative approaches to assessment are concerned with measurement, while qualitative ones are not. The former types are analytic and give rise
Figure 1.1 Purposes of assessment
to numbers of ‘correct’ answers which determine the grade. In this sense they might better be regarded as grading or marking. Rowntree (1987: 10) observed that ‘...“grading” assumes that the teaching is essentially beyond reproach’. Qualitative approaches acknowledge the complexity of learning. They frequently relate to the process of learning, are criterion referenced and approach assessment by means of description and degree of match between objectives and student achievement.

**Process vs product**

What is assessed? Both products and processes of learning may be assessed. Products include traditional academic activities such as examinations and tests, coursework in the form of essays, lab reports, project reports and so on. Processes located within academic settings include class participation, group process analysis, interpersonal, communication or presentation skills development. Processes associated with professional practice include, on the medical side, clinical or dental procedures, practical surgical skills, anaesthesia skills, performance of residents and so on. Teacher performance is another example of a professional practice process. Assessment frequently relates to individual learners, but some of these (particularly processes) also occur in groups (see Hounsell et al., 1996).

---

### Table 1.1 Biggs (1999) quantitative and qualitative approaches to assessment

<table>
<thead>
<tr>
<th>Quantitative assessment</th>
<th>Qualitative assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Learning performances are ‘unitized’.</td>
<td>1 Learning is complexly structured, with new knowledge building on previous.</td>
</tr>
<tr>
<td>2 Assessment is analytic. Marks are allocated to units (sometimes arbitrarily). Units are correct or incorrect.</td>
<td>2 Assessment needs to reflect the complexity of learning structure and is holistic.</td>
</tr>
<tr>
<td>3 Units all ‘worth’ the same. Numbers of correct answers count rather than which are correct.</td>
<td>3 Assessment = grades not quantitative sums. The ‘grade’ is a descriptive statement which notes goodness of match between objectives and student achievement.</td>
</tr>
<tr>
<td>4 Individual performances may be compared.</td>
<td>4 Each student’s assessment is independent of any other.</td>
</tr>
</tbody>
</table>

*Source: Derived from Biggs (1999).*
**Continuous vs terminal assessment**

Rowntree (1987) identified several other conflicting modes of assessment, including continuous vs terminal. Does the assessment take place regularly, or is it limited to the end of module or course? The contrast between coursework and examinations may also be related to this distinction, given that continuous assessment often takes the form of coursework and terminal assessment, the examination.

**Internal vs external**

Where does assessment take place? Is it carried out within the institution by teachers or students, or do individuals external to the institution have responsibility for assessment? Assessment of trainee teachers frequently involves classroom teachers, and, as we shall see later in the book, assessment may now take place in the workplace and involve employers and work supervisors more generally.

**Convergent vs divergent**

This distinction derives from Hudson (1967), who, in his psychological study of the English schoolboy, differentiated two types, the convergers and the divergers. The former, converger, was found to be better at standard intelligence tests than at open-ended tests, while the reverse was true of the diverger. Similarly, assessment can take these two forms. A convergent test is one which has correct or predictable answers, while a divergent test has many possible answers, relying on the creativity and interpretation of the testee.

**Idiographic vs nomothetic**

An *ideographic* methodology, such as that employed in ethnographic and action research, according to Uzzell (2000), captures the richness and complexity of a phenomenon, but risks basing conclusions on a small number of cases which may not represent the population as a whole. Assessments such as diary keeping, portfolio building, Records of Achievement (RoAs) and so on may be thought of as examples of *ideographic* methodologies. Another example is the in-depth interview. Hammond (2000) argued that such an approach may be of great value when the focus of interest is on dynamic processes within individuals. The ideographic approach is often contrasted with *nomothetic* methodology. In this, data are collected ‘by some process of averaging’ (Uzzell, 2000: 326) from a large number of people, and descriptive statistics
such as means and standard deviations, calculated. Generalizations from such data are thought to have some degree of validity. An example of this type of methodology is a questionnaire survey.

**Obtrusive vs unobtrusive**

A final contrast distinguishes between assessment that is obtrusive and that which is not. Is assessment clearly signalled? Are learners aware that they are being assessed? Unobtrusive assessment, though undoubtedly useful in some circumstances, is ethically questionable, as informed consent is absent.

3 What to assess?

Here, we shall review foci of assessments, which can range from the traditional essay, laboratory report and examination to products and processes associated with newer, alternative assessments such as autonomous or ‘authentic’ assessments. We shall focus on traditional assessment in the next chapter, Chapter 2, and learn more about newer alternative assessments in Chapter 3. We shall now review some frequently used, and some not so frequently used, foci.

**Case studies**

A case study, according to Cohen *et al.* (2000: 181), ‘provides a unique example of real people in real situations, enabling readers to understand ideas more clearly than simply presenting them with abstract theories or principles’. Case studies can also bring together theory and practice, establish cause and effect, and help readers recognize the powerful influence of the context on behaviour. Cohen *et al.* (2000: 182) argued that case studies encountered in the assessment literature can be clinical simulations or fictional accounts which are modelled on real life situations. Swanson *et al.* (1995) described the use of an example of the first type, Patient Management Problems (PMP) in the health professions. These are written clinical simulations designed to ‘pose more realistic challenges to medical problem-solving skills’ (Swanson *et al.*, 1995: 5). PMPs begin with a scenario which describes a patient situation. The examinee starts by seeking additional information and then initiates appropriate patient management activities, chosen from a series presented in the simulation. Feedback is then presented about the effects of actions on the ‘patient’.

I have used examples of the second type in both psychology coursework and examination assessments. I prepared fictitious scenarios involving a dysfunctional family. Students were asked to choose any two
members of the family who represent a different problem and answer a number of questions about each. They were required to analyse the family setting and identify factors which may have contributed to the situation, making use of developmental theory and research. Other questions probed stereotypes and required students to suggest helpful courses of action.

Assessing case studies

Case studies may be assessed in both traditional and more innovative ways, and for formative or summative purposes. In both examples cited previously, assessment was carried out by the teacher for summative purposes. Answers to my examination case study question were always interesting to mark, and usually achieved the aim of getting students to apply theory to practice. Marking case study answers shares many of the problems of essay marking. However, much may be achieved by open discussion and agreement between teachers and learners concerning the criteria by which work will be judged. Swanson et al. (1995: 8) noted that scoring ‘rich and interesting’ behaviour can be problematic, and that the variety of responses to this type of exercise makes developing scoring keys difficult. I echo their call for validation work of this type of assessment.

‘Devolved assessment’

‘Devolved’ assessments are those conducted in ‘real’ settings, often by non-academics. Purcell (2001) provided an example of the use of devolved assessments in the context of competence-based learning of accounting technicians in areas where observations in the workplace, witness testimonies and simulations constitute the best sources of evidence. Further details are provided in Chapter 3. I have also used this type of assessment in a work-based module, ‘Volunteering in the community’, in which student volunteers contracted to spend a number of hours each week in an organization of their choice (Falchikov and MacLeod, 1996; MacLeod and Falchikov, 1997). Learning contracts were agreed and signed by each student, an academic and a representative from each organization. Student volunteers were supported by lectures and group activities delivered and organized by myself and a colleague dealing with theoretical aspects of working as part of a group, conflict, decision-making and other relevant topics.

Assessing ‘devolved assessments’

In the volunteering module described above, assessment was carried out in a variety of ways, including some self- and peer assessment and
assessments by the member of staff in the volunteer organization who had responsibility for the student. Our experience led us to conclude that work-place assessors were extremely generous, often failing to differentiate between students, even when two or three were in the same volunteer organization. However, we found good agreement between averaged peer and teacher ratings of oral presentations delivered in the classroom. Other examples of involvement of employers in assessment will be considered later in this chapter. Once again, assessment may be facilitated by the development and use of structures in which outcomes, sources of evidence and criteria are made explicit.

Exhibitions

Exhibitions may be live performances of artistic or technical skill or demonstrations of the products of learning (e.g. Nevo, 1995). They are classed as examples of alternative assessment by Birenbaum (1996) and Dierick and Dochy (2001) and included in a list of methods likely to facilitate self-evaluation and self-development by Klenowski (1995).

Assessing exhibitions

Little is to be found in the literature regarding assessment of exhibitions. However, in Nevo’s (1995) examples, students were actively involved in the assessment process, being required to provide evidence of their ability to use knowledge. Nevo reported that the standards by which students are judged (presumably by teachers) must be stated in advance of the exhibitions. Klenowski (1995) discussed this form of assessment in the context of student self-assessment.

Interviews

‘Interviews enable participants – be they interviewers or interviewees – to discuss their interpretations of the world in which they live, and to express how they regard situations from their own point of view’ (Cohen et al., 2000: 267). Interviews take a variety of forms, and are not always conducted face-to-face. For example, Woodfield and Earl-Novell (2002) reported research which used structured e-mail interviews.

Assessing interviews

Interviews are frequently used in the study of self- and peer assessment, usually in order to evaluate or assess either a person or a group, or to enable the gathering of data about some aspect of the process. However, there is scant information in the literature about how they are used as an
assessment method. Similarly, I have found nothing about how interviews may be assessed.

**Journals/reflective logs/diaries**

Brown *et al.* (1997) argued that learning diaries, logs and journals are essentially the same, in that all are deemed to promote reflection and all are based on the assumption that reflection is beneficial to learning. Some authors, however, differentiate between these activities. For example, Freeman and Lewis (1998) see logs as being less personal than diaries, the former being a factual record of a learner’s activities while diaries are a personal selection of events which can also include some reflection and analysis. Gibbs (1995) places journals between logs and diaries, in that they involve some reflection but are not as personal as a diary. We shall now look briefly at each variant, but consider their assessment together, at the end of the section.

**Journals**

Clandinin and Connelly (1994: 421) saw the purpose of journal writing as to provide a ‘powerful way for individuals to give accounts of their experience’, while Brown *et al.* (1997) ‘unpacked’ the concept of ‘account’, identifying four purposes of learning journals. They argued that this type of assessment tool provides students with opportunities to

- record their learning experiences;
- reflect on progress and problems;
- integrate theories, practice and different approaches;
- express feelings and mood states about their learning.

**Logs**

Brown *et al.* (1997: 196) described a learning log as ‘a personal record of their [students’] experience which they can use to provide a self-report of their . . . learning’. Describing the use of learning logs in the assessment of problem-solving, Brown *et al.* (1997: 151) suggested that logs might include information concerning how problems had been solved, what false starts had been made, ‘what loops in their reasoning occurred’ and what approaches had been found to be helpful. Freeman and Lewis (1998: 254) stressed the importance of making marking criteria clear and explicit. They suggested, as a minimum, that students be told:

- what to record;
- how much detail to record;
the number and range of items to include;
the time period over which to record;
the desired format;
the type of analysis to be carried out prior to submission.

Diaries

Breakwell and Wood (2000: 294) stated that, ‘any data collection strategy which entails getting respondents to provide information linked to a temporal framework is essentially a diary technique’. Diaries most usually involve written records, but can also include photographs, or, more recently multimedia. Diary entries may be specially elicited or spontaneously occurring records. In diaries, although information is usually temporally ordered, frequency of entries can vary, as can the degree of structure supplied or required. Diaries are described as an appropriate form of assessment where the learning is focusing on the processes of doing something rather than on the outcome. Diary writing, according to Ashcroft and Foreman-Peck (1994: 59) can be ‘a way of building up students’ confidence in their own thoughts and feelings, rather than taking on other people’s in an uncritical fashion. It could be said therefore to encourage independent thinking’. These techniques are thought to have a number of advantages in that they are easy to use (though some guidance is still necessary) and they are cost effective. In addition, self-reporting is believed to engender self-revelation, and is, therefore, useful for collecting ‘intimate’ information. However, a number of disadvantages have also been identified. Respondents may under- or over-report events or behaviour depending on perceived disapproval or approval of others. Studies involving this type of assessment can suffer significant drop-out which increases the bias of the sample. In addition, students may be unwilling to participate or may lie. However, problems of unwilling participants, poor recruitment and drop-out can be helped by the formal requirement to keep a journal, diary or log as part of the assessment of a course.

Assessing logs, journals and diaries

As with many of the newer assessment methods, little is recorded in the literature about how logs, journals or diaries may be assessed. Boud and Knights (1994) asserted that setting an assignment which teachers believe will encourage reflection is not sufficient, as the intent of the student is a crucial determinant of what actually happens. Additionally, Brown et al. (1997) pointed out that no study had been done to ascertain whether the use of learning journals changes the style of learning. Based on my own experience of using the learning log method, assessment may be carried
out in a number of ways. Teachers may mark the written record, or portions of it selected by students, as they would an essay, using the guidelines supplied as criteria against which to judge the work. Students may submit a self-assessment of all or part of the record. McNamara and Deane (1995: 21) argued that, in the self-assessment part of their implementation, criteria and ‘means for assessment’ are up to the student. Teacher marking is likely to be summative, whereas student assessment is more likely to be either totally formative or summative only in part. Given the personal nature of diaries, journals or logs, peer assessment is not a suitable method.

Learning contracts

A learning contract has been defined as simply ‘an agreement between a student and a tutor about what and how a student will learn’ (Ashcroft and Foreman-Peck, 1994: 62). In cases where learning contracts are used to assess learning on work placements, the employer is also usually included in the contract. The Maricopa Center for Learning and Instruction Faculty in Progress Program guidelines (Maricopa, FIPP, 2003–04) also make explicit the need for making clear what, exactly, is to be included to support claims that learning has occurred. The guidelines also specify that contracts should be based on personal or professional needs as well as on requirements of the programme. The contract should also include a timeline, including specific dates, months, or semesters indicating when the student plans to complete the activities. Further details are to be found on their website.

The Postgraduate Certificate in Teaching and Learning for lecturers, with which I am associated at Napier University, requires participants in the programme to design their own learning contract for the final module of the course, the Independent Professional Study element. The Napier scheme identifies important features of learning contracts. Learning contracts should

- be written;
- record a negotiated agreement;
- be used to assist the planning of learning;
- used to measure progress;
- put responsibility on the learner;
- give a title;
- indicate a rationale;
- indicate actions that will occur;
- outline how output/results will be presented;
- reveal the criteria for assessment (Earl et al., 2003: 13).
Assessing learning contracts

There seems to be some confusion in the literature regarding what, exactly, is to be assessed in a learning contract. Is it the contract itself or the match between the learning objectives contained within the contract and the evidence supplied by learners? The Napier University Postgraduate Certificate procedures follow both paths. Tutors work with learners to construct a learning contract, which is then scrutinized by a panel of teachers. Learners are not allowed to proceed until the contract and the learning objectives it contains have been approved. Brown et al. (1997) addressed the issue of learning contract assessment in the context of self-assessment of work-based learning. They describe a process, similar to that used by Napier University, in which a contract of learning objectives is drawn up by tutor, mentor and student but which is then later used as a guide to assessment.

Negotiated reporting procedures

Serafini (2000) described negotiated reporting, in the context of compulsory education, as a form of self-assessment based on negotiated criteria which reflect the teacher’s perspective, information from ‘standards documents’, and beliefs and values of the community. Students help create their own report cards using negotiated reporting. The Australian Northern Territory University’s Faculty of Education, Health and Science (NTU, 2003) use a negotiated summary report in the assessment of pre-service teachers. The report is jointly prepared by pre-service teachers and their supervisors. Guidance is given on their website about both structure and content of the report.

Assessing negotiated reporting procedures

Little is to be found, so far, in the literature about either negotiated reporting procedures or their assessment. In the NTU (2003) example, we do not learn the use to which the negotiated report is put, neither are details of the assessment procedure made explicit.

Observations

Cohen et al. (2000) see observation as an attractive technique, because, when carrying it out, you are dealing with a real situation rather than artificial laboratory-based behaviour. However, most situations are complex
and Simpson and Tuson (1995: 22) pointed out that,

One of the most difficult decisions is choosing exactly who or what you are going to select as the observational focus of your study and putting strict boundaries around it so that your study is kept manageable in terms of scale.

Observations can be placed on a continuum from highly structured to unstructured. Cohen et al. (2000) (and, indeed, any good research manual) provide further details for interested readers of how observations may be made.

Simpson and Tuson (1995) reviewed some of the problems to which observational methods can give rise. As they pointed out, observing others at close quarters as they go about their normal business intrudes into people’s personal lives and space, and raises certain ethical issues.

Observation is occasionally used as an assessment strategy in higher education. For example, in Radnor and Shaw’s (1995) ‘reconciliation model’ of assessment, assessment is integrative and formative, no specific instruments are used, and teachers form judgements through observation and students’ responses to set tasks. Similarly, the AAHE (2001a) described a project to be guided by Alverno College faculty, based on the ‘student assessment as learning’ principle. Integral to this process is the observation and judgement of each student’s performance on the basis of explicit criteria. Swanson et al. (1995) reported that observations of students in the workplace, along with witness testimonies and simulations, can provide good sources of evidence of learning, and can constitute ‘devolved assessments’, as we saw above.

Assessing observations

Observation as assessment involves several problems. For example, McMartin et al. (2000) noted its time-consuming nature. Unstructured observations, where no criteria are made explicit and where checklists are not used, are particularly problematic. Structured observations, with operationalized behaviours and explicit criteria, are preferable. There is some information to be found on the Internet relating to the assessment of the observational technique. One of the more useful contributions came from the Center for Applied Research and Educational Improvement at the University of Minnesota (CAREI, 2003). In this, details of an assignment using observational techniques is described in some detail. Learners are provided with task information, hints on how to proceed and a rubric for assessing the final product. It is not clear from the information provided whether assessment of the product is carried
The seven pillars of assessment

out by teacher or student, though the tone of the report suggests the former option.

**Oral presentations**

Oral presentations seem to feature more and more in higher education, and much has been written about their assessment. In oral presentations, students, working alone or in small groups, typically research a topic and present their work to their peers. Several overviews of alternative or new assessments refer to oral presentations as a widely used vehicle (e.g. McDowell, 1995; Birenbaum, 1996; McDowell and Sambell, 1999; Dierick and Dochy, 2001). Reports of individual studies are numerous (e.g. Carter and Neal, 1995). In particular, Magin and colleagues have been involved in studies of assessment of oral presentations at the University of New South Wales for a number of years (e.g. Magin and Reizes, 1995; Magin and Helmore, 1999, 2001; Magin et al., 2001).

**Assessing oral presentations**

Oral presentations are often associated with self- or peer assessment (e.g. Price and Cutler, 1995; Roach, 1999). I have used oral presentations with a variety of students at a variety of levels and for a variety of purposes over a number of years (e.g. Falchikov, 1995a). I find that involving students in the assessment of presentations is extremely beneficial. At its most basic, having to provide feedback to a peer requires concentration, which, in itself makes for more active engagement than simply listening. This is as true for teachers as for students. Other skills may be developed, not least of all the skills of analysis and application of criteria. Tact, too, may be encouraged when having to critique a peer’s presentation.

**Portfolios**

Birenbaum argued that the most quoted definition of ‘portfolio’ is that provided by Arter and Spandel (1992) who described it as a ‘purposeful collection of student work that tells the story of the student’s efforts, progress, or achievement in (a) given area(s)’ (Birenbaum, 1996: 8). Portfolios must include not only the guidelines for selection of items, but must involve students in the selection of portfolio content. They must also contain the criteria for judging merit and evidence of self-reflection on the part of the student. Hall (1992) defines a professional development portfolio for teachers and lecturers as ‘collection of material, made by a professional, that records, and reflects on, key events and processes in that professional’s career’ (Hall, 1992: 81, in Challis, 1999: 370).
D’Aoust (1992) identified three types of portfolio:

- Exemplary folio – best/most representative examples
- Process folio – examples of the developmental process
- Combined folio – 2 sub-folios incorporating exemplary and process elements

Challis (1999) argued that it is difficult to describe a ‘typical’ portfolio, given the very personal nature of this form of assessment. She noted that some portfolios, such as those used simply to record a range of activity, can be relatively unstructured. However, when portfolios are to be used as part of a formal assessment, some structure is desirable.

As we shall see in Chapter 10, Chang (2001: 144) developed and described a web-based learning portfolio (WBLP) system designed to help students produce learning portfolios quickly and easily.

Challis (1999: 370) listed the benefits arising from the use of portfolio-based learning:

- it recognizes and encourages the autonomous and reflective learning that is an integral part of professional education and development;
- it is based on the real experience of the learner, and so enables the consolidation of the connection between theory and practice;
- it allows a range of learning styles to be used according to the preferences of the learner;
- it enables assessment within a framework of transparent and declared criteria and learning objectives;
- it can accommodate evidence of learning from a range of different contexts;
- it provides a process for both formative and summative assessment, based on either personally derived or externally governed learning objectives;
- it provides a model for lifelong learning and continuing professional development.

Keith (1996: 180) also argued that, when used formatively, portfolios become ‘a tool for learning as well as measurement’.

Portfolio-based learning seems to be widely used in the professions. Challis (1999) and Ben-David (2000) reported examples from the medical profession. Challis has also reported examples of portfolio use with pre-registration house officers, in general practice training, with specialist registrars, for continuing professional development and in nursing education. Young (1999) referred to portfolio use in the Health Sciences. Several examples come from teacher education (e.g. Halsall, 1995; Young, 1999; Gosling, 2000). Portfolios are also used in the teaching of English as
a Second Language (ESL) (e.g. McNamara and Deane, 1995). Ashcroft and Foreman-Peck (1994) discuss the use of portfolios in the context of British National Vocational Qualifications (NVQs).

Assessing portfolios

However beneficial a portfolio may be to the user, its assessment may be problematic. Challis (1999: 375) argued that the ‘highly individual nature of each portfolio means that their assessment can present as many challenges as the building of the portfolio itself’. Challis sees self-assessment as a necessary component, but argued that assessment external to the student is also necessary. Ben-David (2000) saw understanding of the criteria involved as crucial in producing agreement between assessors.

Challis (1999) provided guidelines for developing an assessment framework for portfolios. She argued that

- assessment should be carried out within a criterion-referenced framework;
- criteria (benchmarks or standards) should be explicit;
- criteria should link to specific learning outcomes/objectives;
- evidence of learning should be accompanied by explanatory reflective statement;
- evidence must be authentic (relate to learner), appropriate to demonstrate learning claimed (valid) and of sufficient recency for assessor to infer that learning is still current.

Furthermore, Challis (1999: 370) pointed out that when portfolios are formally assessed, we should remember that a portfolio remains the ‘practical and intellectual property of the person who develops it’.

Birenbaum (1996: 11) elaborated a rubric for judging a portfolio as a whole, based on a four point scale. This is summarized in Table 1.2.

Records of achievement and profiles

According to Ashcroft and Foreman-Peck (1994: 63), records of achievement (ROAs, sometimes called profiles) are ‘systematic attempts to involve students in recording, reviewing, and evaluating their academic and personal progress and development throughout their college careers’. ROAs have potential to provide a fuller picture of a learner than a set of examination marks. Various forms of profiling and pupil records have been used in compulsory education for some time (e.g. Broadfoot, 1986) and also in further education (e.g. Mansell, 1986). The UK’s Department for Education and Skills (Dfes, 2003) has a National record of achievement website which, in theory at least, provides information for adult learners,