

# 2

## GETTING STARTED WITH SELF- AND PEER-ASSESSMENT

### ASSESSMENT CRITERIA

#### *Definition and rationale*

The role of the criteria is fundamental to self- and peer-assessment because criteria provide an objective structure for those who generate and implement them. Marking criteria are, in effect, the seat of ownership. Sadler (1989) defines a criterion as 'a distinguishing property or characteristic of any thing, by which its quality can be judged or estimated, or by which a decision or classification may be made'. Without criteria, academics rely covertly on an expert's notion of quality. There is a connoisseurship; judges 'know' what the standards are and how to use them (Sadler, 1989). There is a reliance on tacit assessment knowledge, knowledge regarding assessment that is in the head of the tutor and not necessarily made explicit to students (or other tutors). This approach to marking is not acceptable for the reasons already considered in Chapter 1, primarily, the exclusion of students from the assessment process. Furthermore, Ecclestone (2001) illustrates other considerations for using criteria such as enhanced reliability, as criteria make assessment more amenable to moderation and standardisation.

#### *Self- and peer- assessment: introducing the criteria*

Students are often unfamiliar with marking criteria. Hence, they need to be clearly introduced to them at the beginning of their course of study. Boud (1986) considered primarily self-assessment, but made suggestions which are also applicable to peer-assessment. In order to try to resolve the issue of criteria, students and staff should attempt to clarify the concepts of assessment criteria. Where possible, tutors should not impose, but listen to the student's perception of the criteria. Joint discussion may help avoid any mismatch in interpretation of the criteria or an individual criterion (Orsmond *et al.*, 1996, 1997 and 2000). Further discussion regarding interpretation of the criteria can be found in Case Study 1 (Merry and Orsmond) at the back of this guide. Rust *et al.*, (2003) also investigated student understanding of assessment criteria and the assessment process. They too found a mismatch of interpretation with difficult criteria, such as analysis and evaluation.

Discussions may initially take the form of asking:

'What would be the factors which characterised a good assignment on this course?'

Once this process has been completed, the elements of satisfactory criteria should be considered. This entails such information as:

- the area to be assessed;
- the aims of the assessment;
- the standards to be reached.

#### *Generating assessment criteria*

Boud (1986) highlighted the importance of students reaching their own decisions about the criteria for assessing themselves but stressed the facilitative role the teacher plays in this process. Two techniques for facilitating understanding of criteria are considered:

1. Structured written schedules for developing individual criteria
2. Structured group activities for reaching consensus on common criteria

### 1. Structured written schedules

These provide a list of instructions guiding students through a sequence of steps involving:

- identifying the criteria, which they consider appropriate to apply to their work;
- clarifying these criteria;
- assessing the priority, or emphasis, which they wish to give to each criterion.

Working with students who have perhaps started a task or who are more experienced in self- and peer-assessment might require different types of questions. Brown *et al.*, (1997) gave some suggestions, such as:

- what do I think about what I have been doing?
- how could I improve my approach?

Once satisfactory criteria have been generated in this way, students use them as a yardstick by which they are able to judge their own performance. This might involve:

- awarding themselves a mark with respect to each criterion; and then
- making a statement justifying that mark.

### 2. Structured group activities

Boud (1986) suggested structured group activities if common criteria for a class are required. This will involve the group (groups) identifying, discussing and agreeing upon a common set of criteria. Initially students are briefed that they will be expected to produce a number of criteria. There then follow a number of sessions where criteria are clarified and discussed.

Freeman and Lewis (1998) also provided suggestions for group construction of criteria. These suggestions particularly lend themselves to engaging students with self- or peer-assessment by asking them to consider some of their own existing course-work, perhaps with the accompanying feedback:

- ask students to review their returned assignment in pairs or as individuals;

- ask them to make brief notes, concerning where they gained or lost marks;
- from these notes draw up a list of the criteria students thought the tutor seemed to be using. These criteria can be discussed further for greater clarity.

By viewing student learning as a journey, the ongoing or re-evaluating of the initial joint discussion regarding criteria is necessary to support the learning that the students are undertaking. This can be seen as a first stage in the self- and peer-assessment process.

The above suggestions are helpful, but some care needs to be taken. For example, students who undertake a poster assignment in a level one microbiology session may also be undertaking a poster assignment in level one physiology. The students may think that the end product of the assessment is 'making a poster'. However, the purpose of the assessed assignment may be the demonstration of specific learning outcomes through a poster, not the making of the poster *per se*. So students need to be clear that the physiology assignment requires demonstration of certain processing outcomes, which may, or may not, be very different from those outcomes demonstrated through the creation of their microbiology posters.

The more experienced students become, the more their approach to criteria construction changes. Sivan, (2000) illustrated this well when discussing how students with previous experience of peer-assessment approached criteria construction. Students were seen to take 'a further step and initiated the allocation of different weighting to each criterion and thus were taking even more responsibility for their own learning'. This acceptance of responsibility further develops a student's sense of 'ownership' of the assessment process and further strengthens a student's claim of being an autonomous learner.

The practise and understanding of self- and peer-assessment develops through use as a course or as education progresses, leading to a deeper understanding of the assessment and criteria requirements as the learning journey progresses.

### **Self- and peer-assessment: criteria implementation**

Different types of criteria can be used to generate different forms of judgements. Miller (2003) considered the implementation of criteria within a self- and peer-assessment context. Miller, looking at oral presentation over two consecutive years, wanted to change assessment marking from looking at a few global components of performance such as clarity

and interaction, (used in 2000) to multiple, very discrete components of performance such as 'the presentation included a plan for community and/or work re-integration' (used in 2001). The reason for this shift was to tackle the tendency for markers to assign scores in a very narrow range, concentrating at the high end of the scoring scale. The results indicated that:

- the initial assessment sheet with questions elicited greater feedback from the markers;
- the revised assessment sheet elicited fewer comments and a larger percentage of negative feedback.

Miller explained this result as a consequence of markers being more critically analytical of the presentation when using the revised sheet. However, it could be that, with a set of discrete statements, there was less the students were unclear about, or perhaps the students did not fully understand the statements, and so were unable to structure questions as well. Furthermore, the use of statements may prove limiting or even detrimental to the learning process. This could occur in a number of ways. Firstly, this form of listed 'closed' statements in an assessment sheet may be less than inspiring to students, as the statements could, even if very carefully worded, be based upon the tutor's singular, and perhaps biased, view of what may, or may not, have occurred. Secondly, rather than help students focus on their own personal learning experiences, statements may, instead, severely limit free and honest individual expression from the student, which would then be detrimental to both students and tutors throughout the course, thereby unwittingly limiting the students' learning potential.

Ultimately, Miller (2003) made an important statement, the highly specific assessment instrument, as opposed to the more global instrument, 'produces better quantitative differentiation of levels of performance at the expense of losing qualitative feedback'. It is very important to be aware before implementing specific types of criteria, what exactly you are hoping to achieve with them. Using Miller's example above, are you looking for the 'quantitative' or 'qualitative'? Knowing this not only lends credibility to the assessment process, but provides useful information, particularly for evaluation purposes.

There are some situations where tutors do not always have total control over how a module runs. Tutors are increasingly finding themselves 'team teaching' where the criteria are defined, perhaps by one person, but implemented by the whole team. Tutors, particularly in the case of new tutors, may find

themselves taking over a module where the criteria are already defined. Often in such situations tutors cannot involve students in construction of the criteria, but that does not necessarily mean that students need be excluded from working with those criteria. There are things that can be done. Here is an example:

- Take the marking criteria and consider the terms used to define the aspects of the assignment to be assessed. One example would be if looking into experimental design then part of the criteria may be to consider how 'robust' and 'rigorous' the design is.
- Make a list of the criteria terms used.
- Ask the students to write down their definitions of each term.
- Collect in the definitions, read them and note the correct ones for each term and the other definitions used. This second part is important, because incorrect definitions may give you an insight into student thinking and highlight any common misconceptions.
- Feedback to the students the range of definition they have used, indicating the ones which are correct.

This is not ideal, but at least students do have involvement in the criteria, and furthermore the ground has now been prepared for further discussions, perhaps in subsequent tutorials. Tutorial work is useful for discussing assessment issues. Adams and King (1995) using six tutorial groups engaged students with self-assessment in a class of 120. Race (1998) outlines a process list for peer-assessment which can be used for groups of up to 100 and takes less than an hour to implement.

#### ***Self- and peer-assessment: using marking criteria to help in making judgements***

Making judgements is ultimately what assessment is all about. Having explicit and unambiguous criteria helps this process, but it is still a challenge for students to take their own work and make judgements about it. Peers can be useful in helping students develop their ability to judge. 'While peers may be unwilling to make formal assessment of their peers they may be more positive when students have to give specific feedback of a descriptive nature for the benefit of their peers and no grading has taken place' (Boud, 1986). When using criteria, remember they are reference points in the process of judgement, aids not replacements (Knight and Yorke, 2003).

Making judgements does not, then, simply involve marking work. Peer review is a helpful way to approach peer-assessment (Pond *et al.*, 1995). When asking students to make judgements about their own work or that of their peers, it is important to consider time pressures. For guidance, if considering poster work, students can provide useful written feedback (with or without a mark), on approximately three or four pieces of work in an hour (depending on their complexity).

Elwood and Klenowski (2002) offered students structured support which considered criteria development and help in making judgements about their work, which incorporates a modelling process which considered six separate topics such as, demonstration of understanding of criteria by grading an assignment.

Computer-based programmes are increasingly being used for assessment purposes. The Many Using and Creative Hypermedia system (MUCH) is a multi-user hypermedia tool that supports collaborative learning and has proved to be suitable for peer-assessment (Rushton *et al.*, 1993). Another computer system called 'Peers' has been successfully used to undertake peer-assessment allowing staff and students to determine criteria and weightings for each criterion (Ngu *et al.*, 1995). The case study (Case Study 2) provided by Kuri stresses two strengths of peer-assessment with respect to criteria, empowerment and peer use of feedback. The use of online learning often allows detailed and personalised feedback to be exchanged quickly thus enhancing its potential effectiveness.

Time does need to be allocated to help students in making judgements using criteria as these comments from a study by Brown *et al.*, (1998) illustrate:

'Grading is really hard, to know whether to give them a 2 or a 4, I've know idea how you draw the line, I just know if someone is good or not'.

'I sat there with these numbers and in the end it became a bit random. Perhaps the tutor finds it easier to break it all down into a section, that's up to her. But I just get a general feeling that's all.'

However, the more students undertake exercises involving generating and applying criteria, the more comfortable they become. A quote from a study interview sums this up well:

'When I did it the first time, I need longer to think what grade I should give to this group. Besides, I worried a lot whether I gave a fair mark to others. However, I can do it quite fast this time ... More you do, better you can do it' (Sivan 2000).

Using criteria and making judgements in a

meaningful way does not just happen after one attempt, students need practice to develop the abilities required. Because of this, it is important that, teaching tutors, try their best to get self- and peer-assessment practices implemented at an early stage within a student's university career, and provide a progression of self- and peer-activities throughout a student's university course of studies. Adams and King (1995) give an indication how this could be done. Sluijsman, (2002) presents a framework with guidance on how to realise integrated peer-assessment activities.

However, a note of caution: just because you have worked through a criteria construction process with students, this is no guarantee that all your students will necessarily understand it. Many will, but some may not. The more experienced students become in working with criteria, the fewer the problems of misunderstanding. However, as with any assessment practice some students may misunderstand.

Lack of understanding may be only one cause of disagreements. Hughes and Large (1993) discuss variability in the marking of oral presentation in pharmacology students despite working with agreed marking criteria. They identify issues separate from the criteria such as, how the 'voice' of the speaker may be heard by those near the front of the lecture theatre and inaudible at the back; or how overhead transparencies may be readable at close quarters but unreadable from a distance.

## ASSESSMENT: VALIDITY AND RELIABILITY

Only a valid and reliable assessment processes should be used to determine what learning has occurred. Reliability of assessment is defined by Fry *et al.*, (1999) as 'the assessment process would generate the same results if repeated on another occasion with the same group, or if repeated with another group of similar students'. Validity is defined as 'adequacy and appropriateness of the task/test in relation to the outcomes/objectives of the teaching being assessed', i.e. it measures what it is supposed to measure. These general definitions have been developed; for example, Gielen *et al.*, (2003) consider the validity of assessment scoring, and whether scores are valid. In this respect the criterion of fairness plays an important role.

### *Self- assessment: validity and reliability*

Three studies which consider validity and reliability have implications for implementing and evaluating self-assessment. Boud and Falchikov (1989) reviewed 48 studies of student self-ratings compared to the ratings of students by teachers. Some of the outcomes were:

- comparing results from a number of studies showed that there is no consistent tendency to over or underestimate performance;
- when asked to rate themselves on a marking scale, able and mature students are able to do so. Able students when new to a subject are aware of, and concentrate on, their own deficiencies, and thus underrate their work.

Falchikov and Boud (1989) undertook a meta-analysis of 57 quantitative self-assessment studies used in higher education. Some of the outcomes were:

- the role of seniority or duration of enrolment of the marker was found to be less important than expertise in a given subject;
- explicit criteria led to greater accuracy of rating as did criteria that students felt they owned when compared to criteria that were provided;
- better-designed study was associated with a closer correspondence between student and tutor compared to poorly designed studies.

Boud (1989) raised the question, 'if there is a high correlation between marks generated by students and those generated by staff, why bother with involving students if their contribution makes no difference to the final grade?' He provided two suggestions:

- self-assessment provides practice in the interpretation of the often arbitrary requirements which most public work needs to satisfy.
- expediency: if students can take a greater role in assessment there is the potential for the saving of staff time on the often tedious task of marking.

Regardless of the correlation between marks, considering marks themselves as an important indicator may be missing the point. Topping (2003) comments 'that the high correlation between measures is in any event redundant, and the processes here are at least as important as the actual judgements'. Engagement in self-assessment is a good way to improve performance and nudge students forward in their Zones of Proximal Development. The case study by Rushton (Case Study 3) at the back of this guide stresses the learning that takes place as a result of self- or peer-assessment.

#### ***Peer-assessment: validity and reliability***

Falchikov and Goldfinch (2000) carried out a meta-analysis comparing peer-assessed and teacher marks.

This study can be seen as a companion piece to the paper by Falchikov and Boud (1989). Some of the outcomes of this peer-assessment study were similar to the outcome for the self-assessment study:

- high-quality studies were associated with better peer-faculty agreement than studies of lower quality;
- student familiarity with the ownership of criteria tended to enhance peer-assessment validity.

However, there were differences compared to the self-assessment study:

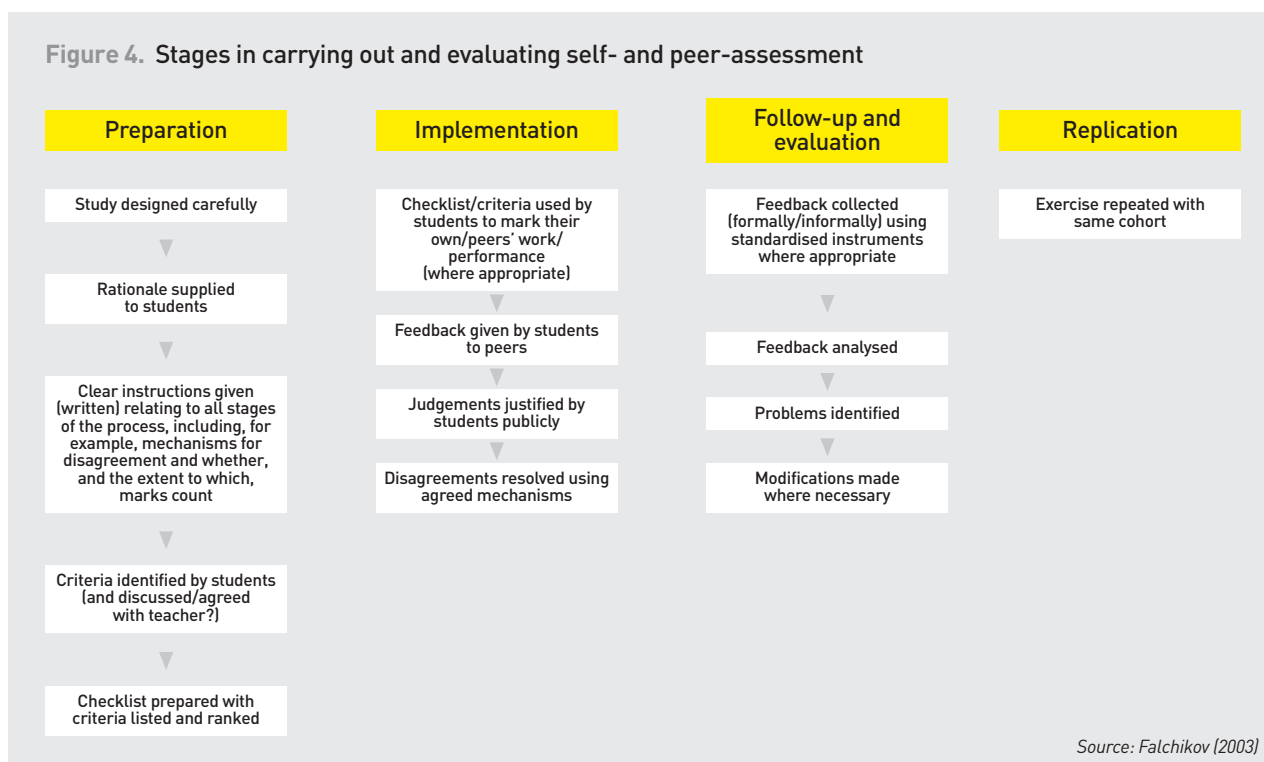
- Unlike self-assessment where the level of the course appeared to be a salient variable, peer-assessment does not seem to be less valid in lower-level courses. A possible explanation is that participants in peer-assessment studies are, in general, well prepared.
- There was no clear subject area difference.

Finally, some areas were measured in peer-assessments that were not considered in self-assessment. Peer marking of several individual dimensions appeared less valid than peer-assessment that required global judgement, based on well-defined and well-understood criteria.

A number of concerns, which have implications for the implementation and evaluation of self- and peer-assessment, have been expressed concerning bias in peer-assessment. Magin (1993), accepting the criticism, where a peer mark is based on an individual peer rating, described a study where multiple ratings are used. The effect of a lenient or severe mark is diluted by the marks from the other students. An interesting approach to dealing with lenient or harsh markers is reported in the case study by Cogdell *et al.*, (Case Study 4 at the back of this guide). In another study, Magin (2001) studied the relational 'reciprocity' effect of peer-assessment of group work. Only 1 per cent of the variance in peer scores was due to bias. In considering gender bias, Falchikov and Magin (1997) used a system that considered student ratings from same and opposite sex, found marginal differences favouring females in peer-assessment.

A brief but useful and in-depth review of validity and reliability in self- and peer-assessment exists (Topping, 2003). Within bioscience, Stefani (1994) found student assessment to be as reliable as that of lecturers, and also reflected on the power/ownership debate, advocating the early introduction in the students' careers of self- and peer-assessment for summative and formative assessment.

Figure 4. Stages in carrying out and evaluating self- and peer-assessment



### SELF- AND PEER-ASSESSMENT: HOW IS IT CARRIED OUT AND EVALUATED?

Developing and implementing the marking criteria are major parts of carrying out self- and peer-assessment. However, there are a number of stages that need planning when carrying out and evaluating self- and peer-assessment. These have been documented (Falchikov 2003). Figure 4 is taken from this work. While this diagram is in some ways self explanatory and a number of issues have been covered already, a little commentary may clarify and refine.

#### Preparation

- Remember the principles of good experimental design. In order to evaluate the procedure subsequently, the dependant variables need to be identified, such as the agreement between peers and tutor, or a measure of the benefits to learning experienced by the participants (Falchikov 2003).
- Students need to be well briefed in advance of the assessment practice and this may mean including details in module/award handbooks, which are often written months in advance of teaching.
- Information should be given both in writing and verbally. Try to ensure that students see and are familiar with the use of all documents, such

as evaluation forms, and have the opportunity to question, clarify and check all material.

- Most importantly, clearly articulate the rationale for using self- and/or peer-assessment. This may make planning the assessment more meaningful. If students have motivation, they are more likely to engage in the assessment task. For example, if the reason is primarily to engage students in the assessment process *per se*, then you will plan differently than if the reason was to make students aware of how to use feedback through peer- or self-formative assessment. Often this is a case of emphasis and the focus of the task.
- Particularly with students new to the process, discuss issues related to fairness and bias (Sivan, 2000).
- Have an effectively detailed approach as to how self- or peer-assessment is to be 'policed' and 'controlled'. Assessment criteria can help with the marking, but some students may under- or over-mark. You need to talk to the students about the consequences of this type of marking (Adams and King, 1995). Race (1998) has some suggestions on how this may be done which include, moderation of peer marking, monitoring student achievement and providing mark-free rehearsal opportunities. Moderation



of marks is a difficult issue as it closely relates to the issues of power discussed in Chapter 1.

### **Implementation**

Perhaps here it may be appropriate to consider when to use self-assessment, when to use peer-assessment, and when to use both. There are no clear rules. However, there are some things you may wish to consider:

- Self-assessment is not undertaken in isolation from others. 'The defining feature of self-assessment is that the individual learner ultimately makes a judgement about what has been learnt, not that others have no input to it' Boud (1995). So, self-assessment should be seen as a continual process, used by students as part of their natural learning. It may initially lend itself to some exercise rather than others, for example PDP. The initials PDP are sometimes assumed to mean Personal Development Plan, but the last P indicates planning. Therefore PDP is a process carried out over time. In PDP students are often asked to reflect and write about those reflections. This is ideally suited to self-assessment or self-evaluation processes. (The distinction between assessment and evaluation is discussed in Chapter 3).
- Sometimes you may wish to use student-centred assessment practices for semi-pragmatic reasons. The case study by Hughes (Case Study 5) illustrates this well. Here peer-assessment is used to reduce marking time, but also provides a range of benefits for students.
- Race (1998) lists a range of activities where peer-assessment can be used, such as in student presentation, interviews and practical work. Peer-assessment also lends itself to group work and there are a number of examples of this (Cheng and Warren, 1999; Freeman, 1995; Goldfinch, 1994; Li, 2001 and Lekj and Wyvill, 2002).
- Assessment practices should relate to the learning outcomes to be achieved. Therefore, check what the outcomes say as this may give guidance as to which student-centred practice to use.
- Boud (1986) gave guidance on giving and receiving feedback. It is a key area and students will need some preparation in matters such as resolving disagreements. To facilitate students in their learning, it is best to introduce guidance early, rather than just before the assessed product is produced, to allow time for students

to assimilate the process.

- Giving students a greater degree of help with marking, and also focusing on the provision of feedback, is important when implementing peer-assessment. The Peer Feedback Marking (PFM) designed to develop peer-assessment was rated by students as conferring more benefits than the more usual lecture marked methods. It also enhanced reflection and the delivery of diplomatic criticism (Falchikov, (1995). The case study by Reed (Case Study 6) gives an indication of how to introduce self- and peer-assessment. Race (1998) also contains some suggestions for getting the most out of peer-assessment such as, allow time for the assessment exercise and the need to keep everyone well informed.

### **Follow-up and evaluation**

Try not to give feedback on the process to the group or individual, but do discuss feedback with students — and receive it, from them. Exchange views with your students as to how the process went. Listen to what value it had for them. Remember, this is an inclusive process. Try not to leave them with a void. Attempt to identify any concerns or problems, which may mean alteration to how self- and peer-assessment may be used next time. Use a detailed and diverse form of evaluation methodology, such as questionnaires and group interviews.

A number of the case studies included in this book have thoughtful guidance on preparing and implementing self- and peer-assessment; Hughes's contribution (Case Study 5) is particularly helpful. The importance of careful consideration of preparation, implementation and evaluation cannot be over emphasised. The consequences of not doing this may lead to ineptly introduced and delivered practices that produce results directly opposite of what is desired.

## **SELF- AND PEER-ASSESSMENT: A COMMUNITY OF PRACTICE**

### ***Encouraging students to engage in self- and peer-assessment***

A barrier to new methods of assessment are an individual's prior experience of being assessed. Hence it is necessary to consider how to encourage students to be involved. Be aware of the problems that students may have. Cheng and Warren, (1997) highlighted some student concerns. Students may:

- be aware of their own shortcomings in the subject area;

- have doubts about their own objectivity; and
- feel that the process is unfair.

Often tutors are regarded as specialists in their field, which could mean that students feel more inadequate as they are novices to the subject and therefore in awe of staff. This should be avoided as much as possible and the role of tutors should be more as guides. This is one reason why a different non-specialist tutor should facilitate in some of the implementation.

(Falchikov, 2003) highlighted other concerns such as.

- Social effects, such as friendships or hostility.
- It is the 'job' of the teacher to mark work.

It may, therefore, be necessary to 'sell' to the students the idea that their involvement in assessment is a good thing. There are a number of considerations to remember.

- Ensure students feel 'safe' about the process. Sullivan and Hall (1997) considered this an important issue.
- In peer-assessment, it may be necessary to make the process as anonymous as possible (Merry and Orsmond, Case Study 1).
- In order to engage students, teachers need to clearly identify why they want students to be involved in the assessment process. Students need to see the value (what's in it for them) of engaging in a particular form of self- or peer-assessment. The case study by Hughes (Case Study 5) has some useful suggestions as to why students need to be involved.
- Show students (and other staff) the research literature as the evidence that it works. Also it may help to introduce students to theories of learning, perhaps discuss with them the Zone of Proximal Development (see Chapter 1) and illustrate how self- and peer-assessment are interventions to enhance their learning.
- Use exemplars as a practice run for the students, so that they can gain objective confidence. Exemplars allow students to understand and use the concepts and criteria with the guidance of their tutor and additional input of peers at the beginning of a project or course. These exemplars may normally be work gen-

erated from previous cohorts of students who undertook a similar assessment. A number of case studies in this guide advocate the use of exemplars, Brennan *et al.* (Case Study 7) is one such example.

### ***Encouraging teachers to engage in offering and using self- and peer-assessment***

It is important to encourage colleagues to be involved in alternative forms of assessment. There are a number of reasons for this.

- Within modular and distance-learning frameworks assessment communities are becoming increasingly fragmented.
- Increasingly, assessment involves more than just one person, or even one subject area.
- This greater involvement will challenge tacit notions of standards shared in a familiar academic community (Ecclestone, 2001).
- The one-off experience is not good. There is a need for practice if skills in assessing their own work or that of peers are to be developed and integrated into students' normal learning patterns.
- Students should have peer-assessment and exemplar work available to support their ongoing self- and peer-assessment practice, enabling and empowering them to achieve higher standards of learning, and therefore higher success in their studies.

Falchikov (2003) gave some suggestions as to how colleagues' suspicions and hostility can be overcome.

- Help allay fears of colleagues by informing them about existing research that advises on best practice.
- Consider using assessment for formative purposes.
- Help ease the change of role required, by stressing the importance of the teacher in setting up, implementing and running a self- or peer-assessment initiative, and in helping students acquire the necessary expertise.



## SUMMARY

Achieving empowerment for students in assessment processes demands their involvement with the assessment marking criteria. Although desirable, it is not always possible to have students involved in criteria construction. However, the onus is on the tutor to ensure that students have a good working understanding of the criteria. The design of the marking criteria often involves discussion of learning outcomes. Therefore it is one of the cornerstones to preparing students for assessment and a valuable tool in successful implementation of self- and peer-assessment. How we 'get started' and 'keep going' with self- and peer-assessment involves a lot of effort, reflection and planning on behalf of tutors. This is well illustrated in the Cogdell *et al.*, case study (Case Study 4). To have a truly effective impact on student learning requires departments and faculties to take on board both the culture and underlying philosophies of self- and peer-assessment; students need to perceive this form of assessment as a natural process in their learning and be actively involved in its implementation and its importance in lifelong learning.