

Beyond feedback: Developing student capability in complex appraisal

D. Royce Sadler

Griffith Institute for Higher Education, Griffith University, Brisbane

Abstract *Giving students detailed feedback about the strengths and weaknesses of their work, with suggestions for improvement, is becoming common practice in higher education. However, for many students feedback seems to have little or no impact, despite the considerable time and effort put into its production. With a view to increasing its effectiveness, extensive theoretical and empirical research has been carried out into its structure, timing and other parameters. For students to be able to apply feedback, they need to understand the meaning of the feedback statements. They also need to identify, with near certainty, the particular aspects of their work that need attention. For these to occur, students must possess critical background knowledge. This article sets out the nature of that knowledge and how students can acquire it. They must appropriate for themselves three fundamental concepts – task compliance, quality, and criteria – and also develop a cache of relevant tacit knowledge. [150]*

Keywords: formative assessment, feedback, qualitative judgment, peer assessment, criteria

Introduction

In the literature on learning research, feedback as a primary element in formative assessment was initially identified with knowledge of results. This referred to simple outcomes (such as a response to a test item) which could be classified as either correct or incorrect. As Kulhavy (1977) put it, feedback is ‘any of the numerous procedures that are used to tell a learner if an instructional response is right or wrong’ (p.211). The implication was that, if learners were told, they would know they should study to remediate the problem. Learning researchers were particularly interested in the effects of various feedback characteristics (especially immediacy, pertinence, data form and type of reward) on the retention of learned material. This focus, which developed out of behaviourist stimulus-response models of learning, is now viewed as unduly narrow.

The feedback characteristics that were important in early studies remain important today, although the focus has changed. In education generally, the emphasis has shifted from promoting the memorisation of factual material to developing students’ abilities to produce responses to assessment tasks that are divergent rather than convergent, and complex rather than simple. Such responses are typically assessed using qualitative judgments. Sadler (1987, 1989) portrayed such judgments as being made directly by person's brain, which therefore functions as both the source and the instrument for each appraisal. Qualitative judgments cannot be reduced to a formal set of procedures which can be applied by non-experts. Broadening the scope of feedback to the point where it promotes complex learning has consequences which are far reaching. Feedback for this purpose may incorporate a variety of elements. These include a symbolic mark or grade to represent the global quality of the work; a detailed explanation or justification of the mark; a description of the quality of the work, with no mark or grade; praise, encouragement or other

affective comments; diagnoses of weaknesses; and suggestions on how to attend to specific deficiencies and strengthen the work as a whole.

It is not feasible here to summarise the findings from the four major reviews of feedback carried out by Natriello (1987), Crooks (1988), Black and Wiliam (1998), and Hattie and Timperley (2007), except to note that all dealt predominantly with school-level learning. It is useful, however, to make a few observations that point towards similar findings in higher education. Feedback is capable of making a difference to learning, but the mere provision of feedback does not necessarily lead to improvement, a fact well known to teachers in all sectors of education, including higher education (Crisp 2007). At the risk of glossing over the complexities of what is known about feedback, the general picture is that the relationship between its form, timing and effectiveness is complex and variable, with no magic formulas. Even something as apparently straightforward as including praise does not necessarily result in learning gains. Butler (1987) found that a crucial difference exists between the effects of praise of the student-self as a person (ego-involving feedback), and praise directed towards how well a task was accomplished (task-involving).

Feedback in higher education

Specifically with respect to higher education, the position generally taken has been that feedback is central to the development of effective learning, partly because assessment procedures play a key role in shaping learning behaviour, and feedback can significantly accelerate that process. Feedback should help the student understand more about the learning goal, more about their own achievement status in relation to that goal, and more about ways to bridge the gap between their current status and the desired status (Sadler 1989). Formative assessment and feedback should therefore empower students to become self-regulated learners (Carless 2006). Obviously, the desirability of feedback cannot be separated from the practical logistics of providing it. Because both effectiveness and efficiency are constant concerns, feedback should not only be of an appropriate type, but also provided within the available resources, especially time for academics to give feedback to individual students. The natural expectation is that students will gain from the feedback to an extent that is commensurate with the effort that goes into producing it. This has led to considerable research directed towards the construction of efficient ways of creating and communicating feedback (Carless 2007; Duncan 2007; Hounsell et al. 2008), and analyses of student responses to feedback (Chanock 2000; Orsmond, Merry and Reiling 2005; Nesbit and Burton 2006; Weaver 2006; Lizzio and Wilson 2008; Poulos and Mahony 2008; Walker 2009).

Although the total number of published research studies specifically focused on feedback in higher education is small by comparison with the number both in other sectors of education and in research on the psychology of learning, the variety is considerable. The studies cited below sketch the background for this article in that they focus on feedback as communication. The emphasis on how students perceive and interpret feedback rather than how teachers intend feedback to be taken is consistent with the broader shift towards student-centred learning. Several researchers have grappled with the problem of how students understand feedback, especially the language in which comments are couched (Higgins, Hartley and Skelton 2001, 2002; Carless 2006; Rae and Cochrane 2008). Among the ways for dealing with the communication problem have been proposals to focus explicitly on raising students' understanding of the assessment criteria in order to initiate the productive dialogues which need to occur (Rust, Price and O'Donovan 2003). Attention has also been directed to contextualisation of the feedback message, the student as receiver of the message, and how the student makes sense of the message. Even the medium of communication seems to make a difference. A study by van den Berg, Admiraal and Pilot (2006) found that 'In their written feedback students commented on structure more than in their oral

feedback. In their oral feedback students commented more on style' (p. 145). This does not necessarily mean that one is superior to the other, only that such differences may well have implications for how students receive and use feedback.

For this article, the point of departure from current research is the premise that, regardless of levels of motivation to learn, students cannot convert feedback statements into actions for improvement without sufficient working knowledge of some fundamental concepts. Teachers who compose feedback obviously possess and draw on a working knowledge which embraces these concepts. They also tend to assume that students are at least adequately equipped as well. However, unless this prerequisite knowledge is identified and addressed, the prospects for even the most thorough feedback are inherently limited.

Scope and terminology

The assessment tasks of specific interest in this article require students to demonstrate higher cognitive skills or forms of professional proficiency. Achievement in these is typically assessed using tasks which require extended complex responses involving analysis, synthesis, creativity, evaluation, or critical thinking. Such responses are common in a wide variety of disciplines, fields and professions at both undergraduate and postgraduate levels. These include languages, humanities, health and social sciences, visual and performing arts, and many areas of science and technology.

The task specifications may be decided by the teacher or examiner, or be negotiated collaboratively with students. In this article, it is assumed that once the specifications have been formalised, they set the design parameters within which students are expected to construct their responses. To that extent, the specifications are meant to be interpreted literally. The assessor is generally referred to as the teacher; in practice it may be a tutor, teaching assistant or other suitably qualified person. Judgment and appraisal are used as synonyms. Feedback is taken to include all communications from a teacher to a student following appraisal of a student response. It may be conveyed in live form such as speech, or in text, audio file or other permanent form that allows for access on demand. An assessment event is concluded when the teacher has evaluated the responses, advised students of the outcomes and provided the relevant information (such as comments, marks or grades) to students and, as appropriate, to the institution. This basic assessment sequence – setting an assessment task, appraising student responses, and providing information about performance – is widespread practice in higher education.

Feedforward

The two main tools for steering students towards producing a work of the type required are the assessment task specifications and, in many cases, information about how the work will be assessed. The actual design is deliberately left open. No assumption need be made about how the information is communicated; the modes may include written directions, verbal or technology-assisted presentations, and face-to-face discussions. All such communications are future oriented and, using Bjorkman's (1972) terminology, constitute 'feedforward'. At this point a caveat is necessary: the legitimacy of providing students with advance notice of the assessment criteria and standards (as a rubric or in some other format) is contested territory. Nicol and Macfarlane (2006), in noting the desirability of prior identification of criteria and standards, cited a number of studies that highlight the tacit component of academic expectations and standards, thereby warning about the inherent limitations of explicit description. Complementary theoretical objections have been elaborated in Sadler (2009b). This issue is not taken further here because it is secondary to the main theme.

Feedback

The first of two broad functions of feedback is to provide a statement of performance through the teacher's assessment of the student response, with or without a rating or grade, and a rationale that indicates how the judgment took into account the strengths and weaknesses of the response. The second function is to provide advice or suggestions as to how a better response could have been prepared. Clearly, if feedback is to have a reasonable prospect of achieving its formative purpose, it has to be both specific (referring, as it necessarily does, to the work just appraised) and general (identifying a broader principle that could be applied to later works). So although feedback is mainly retrospective, it has a prospective orientation as well. This carry-forward tailored component constitutes an additional element of feedforward (Hounsell et al. 2008).

General recommendations in the literature about the desirable properties of feedback include: telling students about the strengths of their works; telling them (gently) about deficiencies, where they occurred, and their nature; telling students what would have improved their works; and pointing them to what could be done next time they complete a related type of response. Throughout, feedback should aim to be constructive and supportive. The volume of feedback for a particular work depends partly on the extent to which the work is deemed salvageable. When a reasonably small number of deficiencies in specific locations would substantially lift the quality if they were attended to, the teacher may provide considerable detail. For really high quality work, there may not be much to be said; for work that is generally weak throughout, the teacher may be at a loss to know where or how to begin.

Teachers who provide extensive written feedback with the recommended properties generally find the exercise labour intensive and cognitively demanding. They give careful thought to precision in wording, because the feedback will later stand as a discrete communication that may be referred to several times. Because the communication is asynchronous, the teacher has to anticipate how the student is likely to react to both the content and tone of the feedback, and this calls for a significant affective outlay on the teacher's part. Alternatives to individually crafted feedback include systematic methods, some of which require only low technology, such as annotating student work with codes that allow students to look up the corresponding feedback statements in a list (Ramsden 2003). They also include electronic templates, rubrics, and criteria sheets, along with databases of feedback sentences from which the assessor can draw to compile extended statements. These can result in substantial savings in time and effort (Case 2007). To the extent that feedback is provided using structured mechanisms or technologies, the labour and cognitive demands can be alleviated. However, regardless of the way in which feedback is produced and communicated, and notwithstanding the limited effect feedback often seems to have, conscientious teachers continue to invest considerable effort in providing it to their students.

Feedforward and feedback as telling

Feedforward and feedback share an important characteristic: as one-way messages from the teacher to the student, they are essentially about telling, or disclosure. Yet despite the teachers' best efforts to make the disclosure full, objective and precise, many students do not understand it appropriately because, as argued below, they are not equipped to decode the statements properly. By implication, teachers who are committed to providing high quality feedback want it to work for their students. Complementary attention should therefore be directed to what students make of the feedback, rather than just its composition. Seen from the learner's perspective, this represents an emphasis on visibility (to the student) rather than disclosure (by the teacher). What does this imply for practice? How can the situation be improved for the student? To start with, those parts of feedback that specifically deal with strengths, weaknesses and especially guidance for improving future works are more than mere conduits of information. For the most part, they are

expository and didactic. The teacher wants the student to learn from the assessment event and the text of the feedback provides the instructional medium. The next question is: How do humans learn from expository text? During the 1960s, this topic was investigated under the banner of 'meaningful verbal learning' by numerous researchers, notable among them being Ausubel (1963) and Carroll (1968).

Obviously, feedback refers to a student work which is, in principle, equally accessible to both teacher and student, so it might be thought that the student's processes of interpreting and learning from it would be straightforward. To test this assumption, it is necessary to take a step back from the feedback itself and analyse the contexts in which, respectively, teachers compose (expository) feedback and students interpret and learn from it. Putting aside any personal biases the teacher may have towards or against particular students, the teacher looks at the work through more or less objective eyes, appraising it as an external observer or consumer. Although the teacher may make assumptions about what the producer was intending, this is not known for certain. It does have an influence, however, on how feedback is framed and guidance given. Additional factors come into play. In particular, the appraisal may be made at a time and site remote from the student (depending on the type of student response), and a teacher may have such large classes that they barely know students except through the works they submit for assessment. The teacher nevertheless accepts a considerable responsibility in trying to turn an assessment episode into a significant learning event.

Students face at least three interpretive challenges in trying to capitalise on feedback. The first relates to the work as a whole: students may focus partly on a work exactly as it was submitted or performed and partly on what they had intended it to be. The learner's personal investment in the production then blurs the boundary between the two. The second challenge arises when the feedback's implications for action are dependent on student understandings of certain concepts or criteria used in the communication. Teachers become accustomed to using certain terms, and can easily presume that students know what they mean. A third type of challenge is experienced by the student who lacks the tacit knowledge (Polanyi 1962) necessary to identify the feature of their work to which some part of the feedback refers. For instance, a teacher may annotate a section with the comment 'This does not follow logically from what goes before'. A student who does not see any problem with the logic cannot take any remedial action. On the other hand, to explain why the logic does not logically follow may require the equivalent of a paragraph or more of explanation, and the teacher either cannot afford the time to compose it, or is unaware that it could be necessary. Either way, the opportunity to learn from the incident lapses.

In all three situations, the student cannot make critical connections between the feedback and the work. The feedback statements then fail as communications, and the telling is to no avail. Recasting the disclosure statements or elaborating them to provide finer detail is futile if the primary referents in it remain fuzzy. Assuming that the interpretation problem is solved, a fourth challenge for learners is to assimilate the teacher's feedback into their existing knowledge bases so that it can be drawn on, as needed, in future constructive activity. As with all learning, newly acquired knowledge needs to be consolidated before it decays if it is to have any positive influence on future works. Both interpretation and assimilation depend on verbalisation and reasoning, and hence familiarity with the concepts and vocabulary that are relevant to translating specifically evaluative discourse into the discourse of (future) production.

The teacher as assessor

The typical teacher's experience with making judgments is now outlined to provide a framework for thinking about how students might develop their personal competence in appraisal. Regardless

of discipline or type of assessment task, teachers may make hundreds of qualitative judgments routinely each year as a normal part of their academic responsibilities (Sadler 1998). This exposes them to a wide variety of ways in which the students argue, describe, compare, evaluate, create, analyse, synthesize and solve problems. They bring the accumulation of judgments about broadly similar responses in the past, and apply this knowledge to new student works. Those appraisal experiences also give access to students' imaginations and strategies, providing vicarious experience about the challenges of production, and extending and enriching the teachers' own knowledge of possible moves which could be passed on to other students. Observe that the teacher's ideas of how a given work could have been done better result from interactions between the teacher's existing evaluative knowledge and actual student works. Sometimes novel, these ideas may not occur to the teacher in the absence of particular student works. This follows simply because students are imaginative and creative too. Also observe that constructive feedback does not need to depend on assumptions about what students may have intended.

The teacher's exposure to a variety of student works gives rise to two distinct types of knowledge. The first is exhibited by the range of overall quality of the entire set of works, extended indirectly by the range of all works the teacher has appraised in the past. This existential range of quality plays a significant role in the formation and maintenance of the teacher's abstract concept of quality itself. It is not unique to responses to any particular assessment task but carries across students and across assessment events.

The second type of knowledge is comparability, which is framed by each subset of student works that are judged to be of about the same quality but are different in their execution. A given level of quality has many possible 'expressions'. This is so much part of the normal experience of appraising multiple works that it is not regarded as remarkable or particularly taxing. The distinction comes as a surprise, however, to many academic teachers at the point of their personal transition from being one among many learners to being an assessor of multiple student works. These two types of knowledge – quality and comparability – can be represented as two dimensions or axes. Each student work can be located at a unique position within the two-dimensional space defined by quality and comparability. In marking a batch of student works, the assessor progressively populates that space with judgments about real cases. It is from within this rich experiential assessment space that the teacher constructs feedback.

The raw material for the feedback derives from the way teachers reach a judgment. As teachers provide explanations for their judgments, they invariably make use of criteria because criteria, by definition, form constitutive elements of evaluative discourses. Furthermore teachers routinely invoke whichever criteria are salient to a particular judgment. This means that they are sensitive to which criteria are relevant and which are not. Behind each criterion sits an enduring quality-related concept that students, too, can acquire. Again, analysing teacher experience provides a clue as to how.

No matter how expertly and conscientiously constructed, it is difficult to comprehend how feedback, regardless of its properties, could be expected to carry the burden of being the primary instrument for improvement. The solution proposed here is to provide learners with appraisal experience that is similar to the teacher's. Desirably, it should be as close in scope and kind as resources will allow. The main tool is, as might be expected, peer assessment (Boud, Cohen and Sampson 2001), for which typologies and various approaches to its implementation are well documented in the literature. However, what is required is not peer assessment as routine activity or busyness, but purposeful peer assessment that is designed with a clear pedagogical intent, namely to provide students with practical experience and a body of conceptual knowledge. The nature of this is now outlined.

Clearly, the gap between the teacher's feedback and the student's appreciation of its practical import has to be reduced or closed. One option could be to teach students the key concepts as a separate activity, outside the context of real assessment events. The alternative outlined below goes back to the aims of higher education, which include having students understand complex concepts and skills relevant to various disciplines, fields and professions. Expressed in basic terms, students can be said to have learned when several conditions are satisfied. The first is that they can do or produce on demand something they could not do or produce before. They also need to be able to accomplish this independently of particular others (such as a tutor or specific group of other students, but not necessarily in strict isolation from others). Finally, the students must be able to deliver the product at an acceptable level of quality. Implicit in this is the requirement that students are able to recognize different levels of quality in works of the types they are being expected to produce. Logically, this is necessary if they are to be able to monitor the quality of their own work while it is still under development (Sadler 1989).

The key feature of the alternative approach is to provide students with substantial evaluative experience not as an extra but as a strategic part of the teaching design. The overall aim is to induct students into sufficient explicit and tacit knowledge of the kind that would enable them to recognize or judge quality when they see it, and also explain their judgments. To the extent that this is achievable, several additional benefits would ensue. These include deeper student engagement with the content and structure of the academic program; a downplaying of teacher-constructed feedback as the critical element for improved learning; and the development of evaluative knowledge and skills of the types that are valued in advanced studies or careers after graduation.

Knowledge of relevant appraisal terms and concepts is necessary for students to be able to think, plan, develop and monitor high-quality works; to be confident in appraising entire complex works; and to engage in intelligent conversations about quality and its determinants. For many students, understanding the key concepts and their implications for practice are non-trivial. Simply being told, even through multiple messages, is rarely effective. As one might expect, a substantial research literature exists on concept formation (also called concept attainment and concept acquisition). The works of Bruner, Goodnow and Austin (1956) and Vygotsky (1986/1934) are seminal in the field, and no attempt is made to summarise their work here. However, for the current purpose, three classes of concepts are identified as important in developing appraisal expertise, namely, task compliance, quality, and criteria. These make up the explicit agenda to be addressed through structured peer assessment. The order in which they are listed here is not accidental but is intended to imply a structural progression. First, the work must be a clear attempt to respond to the task specifications before its quality can be assessed; a judgment can then be made about its quality; and the judgment can be explained or justified with reference to salient criteria. All three are crucial elements in assessing complex student works across many disciplines.

Task compliance

In higher education, the form and structure of what is expected from a student in response to an assessment task is generally set by or agreed with the teacher. As many academic teachers can attest, student productions often do not comply with the basic specifications. Many responses fail to address the specified issue, solve the specified problem, or answer the specified question. Furthermore, the students who submit such responses typically seem oblivious of the mismatch. The issue here is not that students attempt to produce what is required but do it poorly; it is that they produce something else, something that may be related to, but nevertheless sits beside, the

original task. There appears to be little research to date on why this phenomenon occurs, the characteristics of what students produce instead, and what could be done about it.

Despite the paucity of research, however, the problem is widely recognised. Evidence for this is to be found in the numerous sources of advice for students that appear in study skills and assessment guides on higher education web sites and in printed handbooks. In the section on answering examination questions, Race (2007) put it this way: ‘... [M]ake sure that you keep answering the question as written, and not answering what you thought the question asked. The greatest cause of lost marks in exams is candidates wandering off on tangents to the question. All the available marks are for points that are directly relevant to the question. There aren’t any more marks for additional knowledge, wisdom or expertise – however brilliant – that is not required by the question’ (p. 260), and ‘Addressing the question is the art of making your answer an excellent response to the process words in the question. These are important “instruction” words or phrases such as compare, contrast, ... show that... and so on’ (p. 261). For a work which does not clearly address the task specified, a common response by assessors is to appraise the work nevertheless but to deduct marks or credit.

In this section, the approach to tackling this problem is to introduce the concept of task compliance. If students are asked to produce a critique of, say, a public policy, a literary work or a technological solution to a problem, their responses are expected to qualify as critiques. A critique is a form of response that is distinct from an explanation; it involves the student in evaluating something and expressing a considered judgment about it. It is also distinct from criticism, and also from the types of mental activity loosely called critical thinking. To develop a critique, the student must stand back from the content or issue, weigh up relevant features and conditions, examine the evidence or grounds for their appraisal, and then commit the result of an evaluative deliberation to words. A critique has a definite structure to it, which is why it constitutes a distinct class of production. This by no means implies that critiques of the same thing from different students will be the same – or even much alike. Desirably, they will differ. However, critiques have a sufficiently similar purpose and structure for them to be recognisable as falling within the class to which the label ‘critique’ can properly be applied. An explanation, on the other hand, conforms to a different set of characteristics. It is arrived at through different cognitive processes, and results in a distinctive form of final product. Other examples of task types are solution, comparison, proof, demonstration, extrapolation, and scenario.

Task compliance refers to the congruence between the type of response stipulated in the task specifications and the type of response actually submitted by a student. If the specifications call for a critique, a student who submits an explanation, a combination of synopsis and conjecture, or something else altogether is not conforming to the form and structure expected. The production cannot therefore be classified as a member of the class of response required. Its quality *as a critique* cannot therefore be appraised. Giving credit for any work that is not a critique potentially has three negative effects: the student’s action is rewarded and reinforced without requiring a change in thinking and practice; the student gains credit through avoiding a worthwhile educational outcome; and the intent of the assessment task, which is to provide evidence of student capability, is subverted.

Admittedly, not all teachers intend assessment task specifications to be taken literally, and many use familiar forms of words out of habit. This is more likely to occur when academics, in devising assessment tasks, focus primarily on the content they wish to test, paying only peripheral attention to what students should do with that content in order to demonstrate achievement of high-order course objectives. This attitude can carry through to marking: work that shows comprehensive coverage may be rewarded highly, regardless of the task specified. However, to say, ‘This student obviously knows a lot about the topic, and that deserves due recognition’

reflects a lack of awareness of the importance of task compliance, as does its frequent omission from lists of assessment criteria.

Quality

This concept applies in disciplinary and professional contexts where complex learner productions are expected to be non-standardised. The concept of quality in these situations is abstract and typically difficult for students to appropriate for themselves. By quality is meant the degree to which a work comes together as a whole to achieve its intended purpose. When complex phenomena are being evaluated, quality is often determined configurally (Kaplan 1964) rather than as the 'sum' of particular 'measures' of its components. Such holistic judgments may amount to more (but sometimes less) than would result from formal consideration of the various qualities taken separately. These qualities are, of course, usually called criteria (Sadler 2009b). In practice, quality is often easier to recognise when it presents itself than it is to define in the abstract, or account for fully in the particular. Not uncommonly, something significant is lost when attempts are made to express quality in propositional or declarative form, that is, in words, including rubrics and expansions of fixed criteria.

Determinations of quality often require all-things-considered holistic judgments in which multiple criteria are attended to simultaneously in interlaced ways. To make a judgment requires seeing beyond superficial differences in the forms of individual works and into the deeper, subtler and more abstract aspects. It also requires more than taking multiple cross-sections of it, each one tied to a single criterion. This is why quality as a concept is treated here as an integrated property in its own right rather than as somehow 'composed' or built up from judgments made on separate criteria. Students need to be exposed to, and gain experience in making judgments about, a variety of works of different quality, and so populate for themselves the two-dimensional appraisal space outlined earlier. They need planned rather than random exposure to exemplars, and experience in making judgments about quality. They need to create verbalised rationales and accounts of how various works could have been done better. Finally, they need to engage in evaluative conversations with teachers and other students. Together, these three provide the means by which students can develop a concept of quality that is similar in essence to that which the teacher possesses, and in particular to understand what makes for high quality. Although providing these experiences for students may appear to add more layers to the task of teaching, it is possible to organise this approach to peer assessment so that it becomes a powerful strategy for higher education teaching (Sadler 2009a).

Criteria

By a criterion is meant a property or characteristic that is useful in the context of quality and quality determinations. Some criteria (such as word length and referencing conventions for a written piece) are straightforward with sharp boundaries. For these, it is relatively easy to tell when some condition has been met or a rule broken. Other criteria, probably the majority, and including some with disarmingly simple labels, are considerably more abstract. This makes them problematic for students until they become competent users of them, which often requires sophistication and fineness in judgment. These abstract criteria are concepts that do not have sharp boundaries, so they have to become known in the same ways other abstract concepts are formed by individuals and shared in social or professional contexts. Eventually, criteria need to become a regular part of each student's appraisal vocabulary, enabling them to rehearse and to reason through in their own minds as they arrive at judgments, and later to explain and justify these to others.

Coherence (in a piece of academic writing) is an example of an abstract criterion. How well do students understand this concept? Can they recognise low and high levels of it in particular works? Do they effectively recognise this property but use different terminology for it (such as 'linked together')? Can they sense and work towards building coherence into their own productions while construction is under way? The crucial test of whether students understand the (assessment) criterion of coherence is not whether they are able to define it formally. It is whether they can make sound judgments about the coherence of their own works and those of others, and whether they use the term appropriately in explaining judgments about quality, providing feedback for improvement, and conversing with others.

The same could be said of evidence (for assertions), originality or rigour (in an argument), artistry (in a musical performance), elegance (in a mathematical solution or a piece of software), integrity (in a clinical interview), efficiency (in a project design), and cogency (in a legal opinion or policy draft). Students need to understand what these mean and imply for real appraisal decisions. They need to know when particular criteria are appropriate to employ in particular cases. They need to know when to invoke a non-standard, rarely used, criterion because, in a particular case, it is critically important. They need to know when to ignore an ordinarily indispensable criterion because some super-ordinate criterion makes it irrelevant to a particular appraisal. The point being made here is that a deep knowledge of criteria and how to use them properly does not come about through feedback as the primary instructional strategy. Telling can inform and edify only when all the referents – including the meanings and implications of the terms and the structure of the communication – are understood by the students as message recipients.

The sleeper issue with respect to criteria as concepts is that particular terms may mean different things to different teachers, and that these differences remain largely unexplored. Space does not permit a detailed analysis of this, except to note it and make a few brief comments. Although academics teaching different courses might agree on a fixed list of criteria, vigorous debate can follow an attempt to formalise the interpretations of those same criteria in a way that enables consistent use by all teachers and assessors. Furthermore, research studies into how teachers can best convey their expectations (including the meanings of criteria) tend to focus specifically on qualitative assessment within a course rather than the issue of consistency of interpretation across courses, and this is not seen as problematic. Two recent examples are the studies by Andrade (2007) and Rust, Price and O'Donovan (2003) in the USA and UK respectively, where the major interest was in how professors can make their expectations known to students. Consistency in the interpretation of key terms may not pose a problem for individual higher education academics or course teaching teams, but it does for students, when they are faced with interpretations that differ significantly from teacher to teacher and from course to course. Finally, it might be thought that students need to be familiarised with standards as well as the criteria. The reason for not raising this aspect earlier is that standards can be thought of as labels that apply to fixed (but arbitrarily determined) partitions of the quality continuum. The crucial issue for students to grasp first is the constitutive nature of quality itself.

Tacit knowledge

A learning environment that attends to developing students' conceptual understandings of task compliance, quality and criteria will, as a consequence, extend the students' tacit knowledge (Polanyi 1962). It cannot be taught (or caught) any other way. Making numerous and purposeful peer assessments is the crucible within which these three concepts can react and interact. This experiential activity gives rise not just to an appraisal or judgment but also to a body of unseen, unarticulated and often unheralded know-how of the intricate relationships between the appraisal

elements and how they are applied. Competent practitioners and assessors constantly draw on their reservoirs of tacit knowledge; it is the very essence of a great deal of professional expertise.

As an example of this, consider again the practice of competent teachers as assessors. They generally interrogate work with dual agendas operating simultaneously. One of them is to develop a perspective on the work's overall quality, and the other is to take note of particular characteristics (often but not only strengths or deficiencies) that are worthy of special focus or attention. These particular 'noticings' are associated with criteria (often drawn from an undefined pool of potential criteria) which then become useful both in shaping the assessor's emerging perspective, and in constructing feedback comments (Wittgenstein 1974; Sadler 1989, 2009b). The salience of a particular criterion to a particular judgment depends on how the work being appraised is perceived. It is therefore a function of both the condition of mind of the perceiver and the properties of the work being assessed. Which of the potential criteria are singled out as relevant at a particular point has less to do with what is detectable through the senses than with what is deemed to be worth noticing. Knowing what is worth noticing is an essential element of a competent appraiser's tacit knowledge. Only when a student also develops tacit knowledge that corresponds broadly with that of the teacher is the student in a position to make meaning out of the teacher's judgment and its rationale.

Higher education institutions have a responsibility to induct students not only into the mechanics of appraisal, but also into a deep appreciation of how complex qualitative judgments can be made with integrity (Taras 2006; Ellery 2008). This necessitates immersion of both teacher and student in a context of appraisal actions that connect symbolic forms (words, sentences and discourse) with their concrete referents (student works as instantiations), a requirement that applies to evaluations generally (Sadler 1980).

A way forward

The overt aim is to shift the focus away from telling the students about the quality of their work (disclosure) and towards having them see and understand the reasons for quality (visibility), and in the process develop personal capability in making complex judgments. This includes judgments about their own works, both during production and on completion. The key to it lies in educating students in the art of making substantive and comprehensive appraisals in ways similar to those characteristically used by expert assessors. In the implementation of the particular approach to peer assessment described in considerable detail by Sadler (2009a), the raw material about which judgments were made consisted of the works of student peers, which usually supplied sufficient variation in quality and form for the purpose.

There are four basic tasks for peer appraisal, and they can be expressed as questions in the following order. Does a particular response qualify as an attempt to address the issue specified in the task description? This is a category question, not a quality issue, and can be decided only after analysing the work as a whole. The next question is: How well does the work achieve the purpose intended? This gets to the heart of the determination of quality. Third: What are the grounds for the judgment reached, using whatever criteria are appropriate to substantiating the valuation? The counter to the potential objection that not having standardised criteria in advance makes it difficult for an appraisal to be made is that training students into the habit of using preset criteria, criteria sheets and rubrics does not set them up well for lifelong learning, because in the world beyond graduation, hosts of things of all descriptions have to be appraised for a variety of reasons and purposes, and to seek a rubric for each is unrealistic. Added to this is the observation made earlier in this article that the use of preset criteria does not necessarily produce the most authentic judgments. The fourth and final question is: How could the work be substantially improved? This requires advice in terms of the work as a whole, and on specific deficiencies or weaknesses.

This agenda-driven peer assessment becomes a fundamental element in an overall pedagogical approach which can substantially replace traditional methods of working towards mastery of the substantive content of the course. This process would place responsibilities upon the teacher to bring students into a progressively mature capability in making evaluative judgments by providing real but non-threatening practice settings in which they also become calibrated to the norms or quality standards of the relevant discipline or profession. In real life, assessors are constrained by certain conventions. In the context of an academic course, appraisals are bound by the parameters of whatever the task specifications state. Within those parameters, students (as assessors) are able to develop skill in recognising quality as this expresses itself in a variety of forms. Although the recognition of quality is the primary evaluative act, it is not a matter of personal taste. Judgments are made within a defined socio-professional context and need to be substantiated, the typical approach being by reference to criteria that are accepted as salient to each particular case. This general approach draws from the work of Polanyi (1962), and could be described as starting learners on the path towards becoming connoisseurs.

There remain many things that are not known about how best to design assessment events that lead to improved learning for students in higher education. The research studies cited earlier in this article have focused on certain factors that can help or hinder improvement in achievement. Most have been framed and executed within models that depend on feedback as a key stage. The theme of this article has been to propose an alternative model for the organisation of teaching and learning. It challenges the view that feedback from the teacher should be the automatic choice as primary agent for improving learning. Too much attention has been paid at the micro level within the traditional model: what the teacher can do to construct more effective feedback, and what the learner should do to make more use of the feedback provided.

Research into the comparative effectiveness of innovations in the provision of feedback has typically drawn its conclusions from student questionnaires of their experiences and from scores on tests and assessment tasks. However, an average gain in performance when good feedback is given indicates no more than at least some students improve. What of those students who show no improvement? Polling them for their perspectives is unlikely to produce evidence that can lead to further change unless students have had some experience with alternatives. Which of the students who show no gain could fare differently with a different approach, provided it could be implemented with no detriment to those who already know how to interpret and utilise feedback? These are questions for the future.

Conclusion

To sum up, it is commonly assumed that the teacher's responsibility for using assessment events to facilitate improvement in learning is fully discharged by providing extensive verbal communications relating to assessment task specifications (and, arguably, fixed assessment criteria); information about the quality of an appraised work; and advice about how future responses to similar assessment tasks should be tackled. The first of these is commonly referred to as feedforward, the other two as feedback. However, empirical evidence shows that while these communications may be appreciated by students, they often lead to little if any improvement in their subsequent work. Because high quality feedback requires specific input from teachers, and many students seem to take little interest in and benefit from it, teachers can be forgiven for feeling disconcerted or discouraged.

The argument in this article is that the fundamental problem lies less with the quality of feedforward and feedback than with the assumption that telling, even detailed telling, is the most appropriate route to improvement in complex learning. Learning from being told is flawed as a general strategy because the conditions for the statements to make intimate connection with the

student work (with a view to future work) are rarely satisfied. Furthermore, certain forms of disclosure, particularly rubrics and criteria-standards templates, can actually inhibit the formation of a full-bodied concept of quality because they tend to prioritise specific qualities (criteria) rather than quality as a global property. They can also inhibit the development of skill in identifying those criteria that are salient to particular appraisals. Put simply, to depend on telling (as embodied in feedforward and feedback) as the main vehicle for promoting student improvement is to rely on the information transmission model for the development of significant assessment concepts.

Assuming that low student dispositions to learn is not the reason for their failure to capitalise on learning opportunities, the issue is how to create a different learning environment that works effectively. The proposed alternative to the usual sequence is to make intensive use of purposeful peer assessment as a pedagogical strategy, not just for assessment but also for teaching the substantive content of the course. Students need a sound working knowledge of three concepts in particular – task compliance, quality, and criteria – if interactions between teachers and learners are to be formatively effective, and capability in complex appraisal is to be developed. These assessment concepts must be understood not as abstractions but as core concepts that are internalised, operationalised and applied to concrete productions. Without these, key assessment concepts are likely to remain submerged and invisible. To the extent that this model is put into action, the significance of feedback-as-telling is reduced. Of course, if the process were to be entirely successful, the need for substantial reliance on feedback from the teacher would be obviated altogether.

References

- Andrade, H., and Y. Du. 2007. Student responses to criteria-referenced self-assessment. *Assessment and Evaluation in Higher Education* 32: 159-181.
- Ausubel, D. 1963. *The psychology of meaningful verbal learning*. New York: Grune & Stratton.
- Bjorkman, M. 1972. Feedforward and feedback as determiners of knowledge and policy: Notes on a neglected issue. *Scandinavian Journal of Psychology* 13: 152-158.
- Black, P., and D. Wiliam. 1998. Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice* 5: 7-74.
- Boud, D., R. Cohen, and J. Sampson. 2001. *Peer learning in higher education: Learning from and with each other*. London: Routledge.
- Bruner, J.S., J.J. Goodnow, and G.A. Austin. 1956. *A study of thinking*. New York: John Wiley.
- Butler, R. 1987. Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology* 79: 474-482.
- Carless, D. 2006. Differing perceptions in the feedback process. *Studies in Higher Education* 31: 219-33.
- , 2007. Conceptualizing pre-emptive formative assessment. *Assessment in Education: Principles, Policy and Practice* 14: 171-184.
- Carroll, J.B. 1968. On learning from being told. *Educational Psychologist* 5: 4-10.
- Case, S. 2007. Reconfiguring and realigning the assessment feedback processes for an undergraduate criminology degree. *Assessment and Evaluation in Higher Education* 32: 285-299.
- Chanock, K. 2000. Comments on essays: Do students understand what tutors write? *Teaching in Higher Education* 5: 95-105.

- Crisp, B. 2007. Is it worth the effort? How feedback influences students' subsequent submission of assessable work. *Assessment and Evaluation in Higher Education* 32: 571-581.
- Crooks, T.J. 1988. The impact of classroom evaluation practices on students. *Review of Educational Research* 58: 438-481.
- Duncan, N. 2007. Feed-forward: Improving students' use of tutors' comments. *Assessment and Evaluation in Higher Education* 32: 271-83.
- Ellery, K. 2008. Assessment for learning: A case study using feedback effectively in an essay-style test. *Assessment and Evaluation in Higher Education* 33: 421-429.
- Hattie, J., and H. Timperley. 2007. The power of feedback. *Review of Educational Research* 77: 81-112.
- Higgins, R., P. Hartley, and A. Skelton. 2001. Getting the message across: The problem of communicating assessment feedback. *Teaching in Higher Education* 6: 269-274.
- . 2002. The conscientious consumer: Reconsidering the role of assessment feedback in student learning. *Studies in Higher Education* 27: 53-64.
- Hounsell, D., V. McCune, J. Hounsell, and J. Litjens. 2008. The quality of guidance and feedback to students. *Higher Education Research and Development* 27: 55-67.
- Kaplan, A. 1964. *The conduct of inquiry: Methodology for behavioral science*. San Francisco: Chandler.
- Kulhavy, R.W. 1977. Feedback in written instruction. *Review of Educational Research* 47: 211-232.
- Lizzio, A., and K. Wilson. 2008. Feedback on assessment: Students' perceptions of quality and effectiveness. *Assessment and Evaluation in Higher Education* 33: 263-275.
- Natriello, G. 1987. The impact of evaluation processes on students. *Educational Psychologist* 22: 155-175.
- Nesbit, P., and S. Burton. 2006. Student justice perceptions following assignment feedback. *Assessment and Evaluation in Higher Education* 31: 655-670.
- Nicol, D.J., and D. Macfarlane-Dick. 2006. Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education* 31: 199-218.
- Orsmond, P., S. Merry, and K. Reiling. 2005. Biology students' utilization of tutors' formative feedback: A qualitative interview study. *Assessment and Evaluation in Higher Education* 30: 369-86.
- Polanyi, M. 1962. *Personal knowledge: Towards a post-critical philosophy*. London: Routledge & Kegan Paul.
- Poulos, A., and M.J. Mahony. 2008. Effectiveness of feedback: The students' perspective. *Assessment and Evaluation in Higher Education* 33: 143-154.
- Race P. 2007. *How to get a good degree: Making the most of your time at university*. (2nd ed.) Maidenhead, UK: Open University Press & McGraw Hill.
- Rae, A.M., and D.K. Cochrane. 2008. Listening to students: How to make written assessment feedback useful. *Active Learning in Higher Education* 9: 217-230.
- Ramsden, P. 2003. *Learning to teach in higher education*. (2nd ed.) London: RoutledgeFalmer.
- Rust, C., M. Price, and B. O'Donovan. 2003. Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment and Evaluation in Higher Education* 28: 147-164.
- Sadler, D.R. 1980. Conveying the findings of evaluative inquiry. *Educational Evaluation and Policy Analysis* 2(2): 53-57.
- . 1987. Specifying and promulgating achievement standards. *Oxford Review of Education* 13: 191-209.

- , 1989. Formative assessment and the design of instructional systems. *Instructional Science* 18: 119-144.
- , 1998. Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy and Practice* 5: 77-84.
- , 2009a. Transforming holistic assessment and grading into a vehicle for complex learning. In *Assessment, learning and judgement in higher education*, ed. G. Joughin, 49-64. Dordrecht: Springer.
- , 2009b. Indeterminacy in the use of preset criteria for assessment and grading in higher education. *Assessment and Evaluation in Higher Education* 34: 159-179.
- Taras, M. 2006. Do unto others or not: Equity in feedback for undergraduates. *Assessment and Evaluation in Higher Education* 31: 365-377.
- van den Berg, I., W. Admiraal, and A. Pilot. 2006. Designing student peer assessment in higher education: Analysis of written and oral peer feedback. *Teaching in Higher Education* 11: 135-147.
- Vygotsky, L.S. 1986. *Thought and language*. Revised and edited by Kozulin, A. Cambridge, Mass: MIT Press. (Originally published in Russian as *Myshlenie I rech* 1934).
- Walker, M. 2009. An investigation into written comments on assignments: Do students find them usable? *Assessment and Evaluation in Higher Education* 34: 67-78.
- Weaver, M. 2006. Do students value feedback? Students' perception of tutors' written response. *Assessment and Evaluation in Higher Education* 31: 379-94.
- Wittgenstein, L. 1974. *Philosophical investigations*. Trans. Anscombe, G.E.M. Oxford: Basil Blackwell. (Original work 3rd ed. published 1967).