Sadler, D. R. (2013) 'Opening up feedback: Teaching learners to see'. In Merry, S., Price, M., Carless, D., & Taras, M. (Eds.) *Reconceptualising Feedback in Higher Education: developing dialogue with students.* (Ch. 5, 54-63). London: Routledge.

**Chapter summary** Higher education teachers are often frustrated by the modest impact feedback has in improving learning. The status of feedback deserves to be challenged on the grounds that it is essentially about telling. For students to become self-sustaining producers of high quality intellectual and professional goods, they must be equipped to take control of their own learning and performance. How can students become better at monitoring the emerging quality of their work during actual production? Opening up the assessment agenda and liberating the making of judgments from the strictures of preset criteria provide better prospects for developing mature independence in learning.

**D. Royce Sadler** is Senior Assessment Scholar, Teaching and Educational Development Institute, The University of Queensland, Brisbane, Australia

RECONCEPTUALISING FEEDBACK IN HIGHER EDUCATION

DEVELOPING DIALOGUE WITH STUDENTS

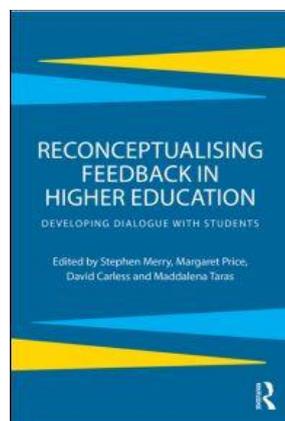Edited by Stephen Merry, Margaret Price, David Carless and Maddalena Taras

Chapter 5

_____

# Opening up feedback

## Teaching learners to see

*D. Royce Sadler*

_____

### The context

Assessment tasks that require students to produce complex responses provide the context for this chapter. Such tasks are typically labelled 'divergent', because no uniquely correct response exists. Examples include: extended solutions to non-routine problems; literary, artistic, musical or other creative performances and works; critical and analytical investigations and reviews; theories and abstract syntheses; computer software; technological, engineering and architectural plans; industrial and product designs; medical and other health-related procedures; project reports; social and other policies; financial, administrative and management schemes; and sophisticated models of interactive systems. Responses may require high levels of technical or procedural skill; high levels of intellectualization, cognitive activity or aesthetic appreciation; or all of these. Some require special abilities that are not easily categorized.

   Graduated complex tasks provide settings and practice sites for students to develop their knowledge and skills. Student responses provide the bases for inferring the depth of their learning. The quality of each response is judged by making a refined qualitative appraisal using multiple criteria. After graduation, complex works typically serve specific professional or discipline-related purposes or ends. Because the focus here is on complex works in general rather than works in specific disciplines, fields or professions, the following terms are used interchangeably: achievement, attainment, performance, accomplishment, capability, proficiency and competence.

### Learning for complex outcomes

Three basic requirements for learners to become proficient in a given domain are that: they acquire a concept of high quality and can recognize it when they see it; they can with considerable accuracy judge the quality of their works-in-progress and connect this overall appraisal with particular weaknesses and strengths; and they can choose from their own inventories of potential moves those that merit further exploration for improving quality (Sadler 1989). Judging

one's own work accurately and dispassionately requires substantial personal detachment from it. Furthermore, in order to develop and become competent and confident in self-assessment, students need experience in doing it independently of their teachers and peers.

The task of teachers is not to coach students through the production of particular complex responses by offering ongoing judgements about quality together with advice on how to improve them. It is to teach students how to judge quality and modify their own work during production. In practice, this involves attending simultaneously to dual agendas: the large scale (how the work is coming together as a whole) and the small scale (aspects that require further attention). For all of this, learners need to have acquired the relevant skills, know-how, and knowledge. Developing this capability normally requires repeated practice on a range of tasks and, in many cases, considerable experimentation to and fro during the work's production. In developing an essay, for example, students may revise their work many times. Even experienced academics often do the same when writing journal articles. However, in certain professions practitioners must attain a very high level of proficiency so they can perform given procedures correctly the first time; recovery or salvage may be impossible afterwards. In such cases, learning with practice typically takes place in non-critical situations first.

Key questions are: How do accomplished producers engage, cognitively and procedurally, in consistently creating complex works of high quality? What knowledge and skills are necessary and how are these applied in practice? The design of learning environments needs to be influenced by answers to questions such as these if students are to eventually master what is required. Students can then carry that capability forward to new situations in more advanced courses or outside formal education. In all of this, where does feedback fit in?

Feedback is often regarded as the most critical element in enabling learning from an assessment event. In practice, it often seems to have no or minimal effect. Whenever creating good feedback is resource intensive, this produces a low return on investment. What can be done? Merely improving the quality of feedback and how it is communicated to learners may not be enough. The proposition argued in Sadler (2010) is that the major problem with feedback is that it has been to date, and is virtually by definition, largely about telling. Research into human learning shows there is only so much a person typically learns purely from being told. Most parents know that already. Put bluntly, too much contemporary assessment practice is focused on communicating better with students.

Teaching by telling is commonly dubbed the transmission model of teaching. It portrays teachers as repositories of knowledge, the act of teaching being to dispense, communicate or 'impart' knowledge for students to learn. Consistent with that was an early conception of feedback as 'knowledge of results' – simply telling students whether their responses to test items were correct or not. Telling is limited in what it can accomplish unless certain key conditions (touched upon

_____

later) are satisfied. By itself, it is inadequate for complex learning. Being able to use, apply, and adapt knowledge, or to use it to create new knowledge, requires more than merely absorbing information and reproducing it on demand. Educational research over the past few decades has recognized the need for students to interact with information and skills, to make these their own, to incorporate them into their existing knowledge bases and structures, and to 'construct' or build knowledge that will serve them as adaptable personal capital. This emphasis may be relatively recent, but the process itself is what humans have always done. What has changed is the perspective.

To the extent that the traditional feedback model operates under transmission-model thinking, it reinforces the assessor's role in providing judgements, explanations and advice as new information (formal knowledge) for students to appropriate. In so doing, it largely misses the point. The knowledge base of learners has to be appropriately organized and sufficiently extensive for learners to be able to convert feedback into deeper knowledge and, ultimately, improved proficiency. If a student's knowledge foundation is deficient, personal appropriation of feedback simply cannot occur. Evidence for this comes from research on the development of expertise (Bereiter and Scardamalia 1993). An important goal should be to avoid treating feedback as a homogeneous commodity conveyed from assessors to learners, one-way telling being the main vehicle. Activities involving peer assessment (Bond *et at.* 1999) and recent work on dialogic feedback by Nicol (2010) represent substantial improvements on that. However, the focus needs to shift away from the narrow issue of how feedback can be improved and communicated, and towards the wider issue of how *assessment* (rather than feedback) can enhance student learning. The problem can then be framed differently. Any assumption that feedback must remain the primary assessment related tool inhibits opening up the agenda.

## Competent producers of complex works

For many complex works, no precise blueprints exist. There are no algorithms or production formulas for getting to the end product, and no optimal path to it. Indeed, the nature of the end product is not necessarily known beforehand. However, during productive activity or the process of solution, its nature develops or is clarified as it emerges progressively. Producers may start with what they think are more or less clear ideas of the problem and only tentative ideas of solutions or strategies. As their work proceeds, both the full character of the problem and potential ways to solve it take concrete form as trial moves are made and their effects analyzed. Amendments are made and, at various stages, overall designs modified, the whole activity being directed towards the development of works of consistently high quality. Producers know when and how to adjust provisional plans and which alternative moves to try next when some do not work out. They notice things that matter and must be attended to, and pay no or relatively little attention to minor aspects that do not matter. They

recognize why and how the organisation of each work may need to be revised as it starts to take shape. As all this is going on, they may - or may not - verbalize and reason out parts of it.

Depending on the nature of the end product, competent practitioners can identify features which require protection, elaboration or relocation as necessary. The knowledge and thinking of producers grows as they engage in sophisticated contingency management. They perceive certain aspects as salient in the circumstances and know to make local adjustments for the benefit of the overall quality or effectiveness of the work. They identify their own mistakes, recognizing weakness or incompleteness. They know that some changes will necessitate compensatory moves or decisions elsewhere, and draw elements and tactics from their repertoires to make them. They may find it difficult to explain to themselves or to others the reasons for all their actions, but they know when changes are needed. Throughout, they project themselves into specific situations: the perspective or position of the subjects of their attention (in certain human service fields and professions); or into the role of audience, consumer or observer (if they produce live or artefactual works). In short, they possess a refined sensitivity to contextual cues as these occur on the run, including shortcomings of their own creation.

## The concept of 'knowing to'

To the extent that this is a reasonable portrayal of what competent producers do, the issue for academic teachers is to figure out how learners can acquire the types of competence required. As indicated earlier, feedback as traditionally conceptualized largely involves the assessor running with dual agendas (overall quality and matters of detail) but the most problematic aspect of the classical feedback model is that assessors (whether academics or student peers) are the ones who do the noticing, the thinking about repair and modification, and the generation of ways to improve. Learners need to develop awareness and responsiveness so they can detect anomalies or problems for themselves. They need to know when something matters in and of itself, and when exactly the same thing matters in one context but not at all in another. This is a significant contextualized figure—ground skill that must be developed. It constitutes a distinct form of knowing, 'knowing to', which involves detecting, 'sensing' without necessarily using the standard five senses. It begins with a feeling of unease or discomfort that something is not as it could be, and that a change would improve correctness, efficiency, flow or elegance. This type of knowledge cannot necessarily be made explicit, that is, expressed in words. It nevertheless exists, even when a person cannot define it in concrete terms or otherwise explain it.

'Knowing to' accounts for some of what Polanyi (1962) called 'tacit knowing'. Wittgenstein (1953: XI, 93) observed: 'I contemplate a face, and then suddenly notice its likeness to another. I see that it has not changed; and yet I see it differently. I call this experience "noticing an aspect"'. Abercrombie (1969) in

_____

her seminal work on judgement captured some of the same essence in her discussion of the influence and intricacies of perception and prior expectations on what is noticed and deemed to count as data in a particular context. Consistent with the work of Polanyi, Wittgenstein and Abercrombie is that of Dreyfus and Dreyfus (1984: 225): '[E]xperience-based similarity recognition produces the deep situational understanding of the proficient performer.' Simultaneously with the apprehension or understanding of a complex situation, the 'associated decision, action or tactic presents itself.'

The knowledge and skills students require do not, and cannot, come about by being told about them or by any form of explicit teaching. How can one person explain to another what they should notice in a particular context (including those of their own making) but is not worth noticing in other contexts if the possible occurrence or existence is not known in advance, and a multitude of things (features, aspects, properties, characteristics) are at least potentially worth noticing? Then, once something is noticed, knowing to do something or at least try. Learners must develop know-to knowledge directly through experience if they are eventually to become self-monitoring.

Educators can provide open assessment environments in which learners grasp the subtle skills required, at the same time ensuring that the process does not become unduly labour intensive for either party. Such environments immerse learners in decision spaces that are similar to those inhabited by the already competent, the closest of whom is presumably the teacher. It involves understanding the macro-level determinants of quality through taking evaluative action in contexts where configural (rather than componential) judgements are demanded. Equally, it involves understanding the micro-level determinants and knowing how to shape the work as a whole through small-scale tactics. This type of 'seeing' typically goes unrecognized in most of the research on assessment for learning, where the focus has been on feedback. It is not esoteric seeing, because perceptive and selective noticing are everyday activities.

The substantial literature on the nature of expertise and how it is developed is an important resource for further thinking. As well as the authors listed above, this literature includes contributions relating to the development of competence by medical and health practitioners, airline pilots and many other professionals involved in complex decision contexts. It turns out that a great deal of what experts do cannot be expressed completely in words and thus codified; it must be learned by experience. Progress on an alternative strategy can be made by seeking answers to questions such as these: What do competent producers do? How does their expertise manifest itself? How did they get that way? How can higher education teachers enable students to develop that type of capability? How can students be launched on trajectories towards sophisticated professional capability by the time they graduate? This is an important thrust of Bereiter and Scardamalia's (1993) research.

Another question remains: Are there any circumstances in which external feedback regularly leads to improvement? The answer is emphatically affirmative,

but only when certain conditions are met. The most crucial condition is that the producer already possesses enough explicit and tacit knowledge to understand the full implications of the feedback. Only then can appropriate action be taken. To illustrate, consider manuscripts being evaluated for publication. Some authors do not know enough about academic writing and publishing for the reviewers' feedback to be understood and utilized except as a recipe to be followed blindly. The feedback is situated outside their 'zone of proximal development' (Vygotsky 1986). Other authors understand that something they happened not to initially 'see' when they produced their work has been subsequently 'seen' or 'noticed' by a reviewer. Its significance is then grasped immediately. External feedback can be a powerful tool for improvement, but can have this impact only when the requisite knowledge base has been established.

## Case account

This account is of an approach to inducting undergraduate students into the types of processes outlined above and represents but one way to provide students with appropriate evaluative experience (Sadler 1989). I have used a similar process in professional development for academics. It is based on my own attempts, but includes some refinements I now recommend in retrospect. The aim was to provide an evaluative environment in which students could develop their ability to:

• make realistic, honest holistic judgements focused exclusively on the quality of academic works similar to those they produce themselves;
• distinguish between aspects of works that were germane to the judgement and pass over aspects that were routine, normally expected and not deserving of special mention;
• construct sound rationales for their judgements; and
• develop their personal repertoires of tactics and moves for improvement.

An additional aim was to provide students with a sufficiently positive experience of learning to motivate them to participate fully, even though none of the work produced would be credited towards the course grade.

The approach adopted was a form of peer assessment with a specific agenda in mind. Inter-scorer reliability was put to one side as secondary to the main exercise. For each tutorial, students had to create a short work of about 300 words in response to a specified academic task and bring three identical de-identified copies to the session. High quality responses required substantial cognitive activity and engagement to address a novel and previously unseen issue. Students had to distil, process, apply and integrate material from different sources (course lectures, notes, websites, textbooks and discussions) rather than reproduce, adapt or compile existing content.

_____

Tutorial groups, each of about 20 students, met with me for one-hour sessions. The activities replaced the tutorial exercises and discussions used previously. Close engagement with the new processes became a primary pedagogical approach, and were billed as such to the students. In the first round of a tutorial session, one copy of each student response was put into a pool and then randomly allocated to another student. Two or three appraisals could be completed in a one-hour session. Students without completed 300-word papers would have nothing to do during the tutorials. As it turned out, attendance remained high.

The evaluative exercise ran along the following lines. The first step was to make a judgement: 'What is the quality of the work you have just received to appraise?' Students' initial concerns followed a fairly consistent pattern: 'Where are the criteria? How can I make an appraisal without criteria or a rubric?' Most of the students' earlier peer assessment activities had employed preset criteria. My response was that they should look closely at the quality of the work, think hard and analyze their reactions. I urged them to be patient, despite their sense of insecurity, and reiterated: 'The aim here is to learn about overall quality, to recognize it when you see it. What is the quality of the work you have before you? What do you react or respond to as you read it? How does it come over?'

Initially I had asked students to record their judgements on a numerical scale, because I wanted them to commit to a definite representation of their considered holistic judgement. However, some students put as much or more effort into deciding the number as into making the judgement. This clouded the issue, as would the use of any familiar symbols or scales. Interpretations of marks, letters or grade division points are invariably loaded towards past marking practices and habits. For example, a mark of 75 on a 100-point scale may have traditionally been the cut-off for a particular designation of quality (an A, or First Class). I now recommend that students represent their judgement of overall quality by the placement of X on a fixed-length line segment without scale points of any kind, 'low quality' being on the left and 'excellent' on the right.

The second step required students to justify their appraisals in 50 words or fewer, sticking to the qualities and properties of the work itself and avoiding praise or censure. Students necessarily invoked whatever criteria were necessary to explain their judgements. This step provided raw material for analysis in later whole-group discussion: the criteria invoked, the terms used, and the necessity to make judgements and provide rationales that would be tailored to each work individually but in all probability be different from one work to another. The third step was to provide written advice to the author as to how either the appraised work could be improved, or future similar works made better.

The concreteness of the three steps was important. People often think they know something thoroughly in the abstract. However, expressing it in material form and thus externalizing it makes the structure of the 'held' knowledge fuller and tidier. Furthermore, a certain seriousness of purpose enters into the exercise from knowing that an appraisal, its justification and suggestions for improvement

will be later accessed by the producer of the work, especially if the activity design includes discussion (Abercrombie 1969).

On several occasions, students observed that the work being appraised did not actually address the specified issue at all. It consisted simply of material obviously garnered from different sources, technically relevant to the subject-matter mentioned in the task specifications, but not directed towards addressing the issue, solving the problem or answering the question. How could its overall quality be assessed if it was not what it was supposed to be? These students had made clear distinctions between the subject-matter content itself, and operations to be performed on that content to address the issue specified. Teachers commonly encounter this phenomenon, but many students do not recognize this as a potential problem. Yet this recognition is crucial for students who have not previously made connections between a task as it is specified, their submitted responses, and the feedback they subsequently receive.

As would be expected, not all students discovered this for themselves. I now recommend that a fourth step be added: 'Did the response actually attempt to address the issue stated?' This should be implemented only after students have had some experience with the original three steps to allow room for individual discovery. However, it should then become the first step to be attempted, because it is pre-emptive: unless the student work can be recognized as an attempt to address the task as it is specified, the question of how well it does so makes no sense. I also recommend that student authors and critics engage in forthright discussion about the works they appraise without becoming sensitive or defensive. My goal was to provide students with practice at being objective, which is why only the characteristics of each actual work were relevant.

As to the quality of feedback provided by other students, it could range from strong and on target to weak and misguided. An explicit point was made with students that this could happen, and that part of their responsibility was to learn to evaluate the quality of feedback. However, if students received essentially the same message from various sources, they should take it seriously. A related issue was this: What should be done if the overall level of student appraisal capability is low, so that they simply share their ignorance or inexperience? In each randomized pool of responses I included my own de-identified attempt at the same task (which also provided me with a check on the adequacy of my task description). In exchange, I received one student response that I appraised exactly as the students did. By the end of each session, at least some students in each round became aware of what my high-quality response looked like. But I had also received feedback on my work from several students. This informed me about their interpretations and misunderstandings. Putting my own work into the pool helped to allay fears and promote trust. After several sessions, I distributed to every student a copy of my own attempt at the task then current, they assessed it, provided their rationales and we engaged in group discussion. These activities constituted my attempt to calibrate students against a level of quality that was not limited to the ideas of students within a particular tutorial group.

_____

Two key features of the assessment environment outlined were deliberate. First, students mutually appraised multiple works that were all responses to the same assessment task. As explained in Sadler (2010), for students to develop a concept of quality, they need to see as extensive a range of quality as possible, and also to see and appreciate how quite different works can be legitimately judged to be of about the same quality. Second, the identification of criteria followed (rather than led) the making of a judgement so that the role of criteria would remain important but subordinated to the main task of holistic appraisal (Sadler 2009a, 2009b). Peer assessment protocols that employ the same criteria for all works are foreign to this process. That somewhat different criteria are typically invoked for different works, including those of the same quality, is not an aberration; it is precisely the point.

Taken together, the four steps require students to focus intensively on the characteristics of the actual works as they stand, to look – and see – without the strictures of any standardized appraisal template. The main intended benefit was for students as budding assessors, not as constructors of feedback for peers or as consumers of feedback from peers. Students need to develop the practical ability to make both macro and micro appraisals, initially of works other than their own. This should improve the likelihood of later transfer and application of that knowledge to self-monitoring the quality of their own works during production.

## Conclusion

Much more than we give credit for, students can recognize, or learn to recognize, both big picture quality and individual features that contribute to or detract from it. They can decompose judgements and provide (generally) sound reasons for them. That is the foundation platform for learning from an assessment event, not the assumption that students learn best from being told. They need to learn to discover what quality looks and feels like situationally. They need to understand what constitutes quality generally, and specifically for particular works. Equally, students need to be able to detect aspects that affect overall quality, whether large or small, and understand how and why they interact. Students need a vocabulary for expressing and communicating both what they find and how they judge, at the least for that part of their evaluative knowledge they can express in words. Only after students have acquired a sufficient basis of appropriate tacit knowledge can they can understand the content and implications of a marker's feedback. At that point, feedback can be effective as learners become more discerning, more intuitive, more analytical, and generally more able to create, independently, productions of high quality on demand.

_____

## References

Abercrombie, M.L.J. (1969) *The Anatomy of Judgement: an investigation into the processes of perception and reasoning*, Harmondsworth, Middlesex: Penguin.

Bereiter, C. and Scardamalia, M. (1993) *Surpassing Ourselves: an inquiry into the nature and implications of expertise*, Chicago: Open Court.

Boud, D., Cohen, R. and Sampson, J. (1999) 'Peer learning and assessment', *Assessment and Evaluation in Higher Education* 24: 413-26.

Dreyfus, H.L. and Dreyfus, S.E. (1984) 'From Socrates to expert systems: the limits of calculativc rationality', *Technology in Society* 6: 217-33.

Nicol, D. (2010) 'From monologue to dialogue: improving written feedback processes in mass higher education'. *Assessment and Evaluation in Higher Education* 35: 501-17.

Polanyi, M. (1962) *Personal Knowledge*, London: Routledge & Kegan Paul.

Sadler, D.R. (1989) 'Formative assessment and the design of instructional systems', *Instructional Science* 18: 119-44.

—— (2009a) 'Indeterminacy in the use of preset criteria for assessment and grading in higher education', *Assessment and Evaluation in Higher Education* 34: 159-79.

—— (2009b) 'Transforming holistic assessment and grading into a vehicle for complex learning', in G.Joughin (ed.) *Assessment, Learning and Judgement in Higher Education*, Dordrecht: Springer 4: 45-63.

—— (2010) 'Beyond feedback: developing student capability in complex appraisal, *Assessment and Evaluation in Higher Education* 35: 535-50.

Vygotsky, L.S. (1986) *Thought and Language*, revised and edited by A. Kozulin, Cambridge, Mass: MIT Press. (Originally published in Russian as *Myshlenie i rech*, 1934.)

Wittgenstein, L. (1953) *Philosophical Investigations*, translated by G.E.M. Anscombe, Oxford: Basil Blackwell.