The respective roles of formative and summative assessment: How to get the best of both worlds

D. Royce Sadler
Griffith Institute for Higher Education, Griffith University, Brisbane. r.sadler@griffith.edu.au

Presenter's prompts for a presentation to the Engineering Faculty, University of Alberta, Edmonton

Part 2

I now want to change tack and explore with you some conditions for assessment to be formative, and the **limited role that feedback can play** in this. It is a fallacy that feedback is *the* critical element! Feedback's potential sufficiency for bringing about improvement is seriously overrated.

- Common recommendations about the structure of feedback:
  - Tell students **what is good** about their work;
  - Tell them **what's wrong,** and where;
  - Tell them what they **could have done** to have made it better (past tense); and
  - Tell them what they **should do** next time in order to produce a quality work (future tense).
  - Be **constructive and supportive** throughout.
- Feedback as commonly practiced:
  - Is often **labour intensive**;
  - Requires a **lot of thought** to get the wording right; there is an **emotional investment** in it;
  - Has to be both **specific *and* general**; and
  - Often **does not lead to improvement**. (What goes wrong?)
- Proposition: The problem is precisely and fundamentally because the dominant way of conceptualising and 'doing' feedback is **based on telling**.
- Returning to the three basic requirements for improvement in an intelligent rather than random way (listed in Part 1). The learner has to possess a concept of **high quality as a goal**, or reference level, to be aimed for; c**ompare** the actual (or current) level of performance with the reference level (appraisal, judgment); and engage in appropriate action which leads to some **closure of the gap**.
  - The (macro) process of grading involves the first two in that it is essentially comparing a particular case either with a reference level or with one or more other cases.
  - Control during production, which is the site where students operate most intensively, is a (micro) process carried out in real time. It involves all three conditions, which must be satisfied simultaneously, not as sequential steps.
- Obviously, students can monitor the emerging quality of their productions – and make adjustments as necessary – **only** while they are actually engaged in doing it.
  - They must attend to the *large scale* (how the work is coming together as a whole).
  - And they must attend to the *small scale* (point by point tactical decisions).
  - For both of these, learners need to have acquired the relevant skills, know-how, knowledge.
- Focus now on the goal, the quality being aimed for.
  - Quality as a concept is like many other abstract concepts (democracy, religion, integrity, elasticity, energy, family, syndrome, justice, coherence).
  - For some simple concepts, drawing attention to properties using positive and negative instances (with appropriate labels) may be enough. This often works well with children.
  - For sophisticated concepts, that is hardly ever enough. Maybe it is never enough.

- Roller coaster example. How would you convey to a person who has no idea what a roller coaster is, and what it is like to ride on one? How would convey the curious mixture of acceleration, exhilaration, fear, surprise, sense of 'capture' by the safety equipment, and lack of control over anything?
    - Describing? Illustrating? Show the person a real one? Show the person a number of different roller coasters?
- Coming back to knowledge about quality requires three elements:
    - Description
    - Exemplars
    - Tacit knowledge (which is partly the background knowledge of how the descriptions apply to the cases, but mostly knowledge of the overall constitutive nature of quality as an abstract (non-concrete) concept.
- Observations
    - None of these three is expendable. All three must be connected.
    - How well these are incorporated into the learning context is basically under the control of the sender of information (teacher and institution).
- How do these depend on the characteristics of the receiver of the information? The **development of tacit knowledge is way underplayed**. Conveying fullest message requires:

  [Shared access to common exemplars] + [Shared language] + [Decisions and explanations]
    - Of these it is the **common experience**, taken literally, that lags seriously behind the others.
- How does the fullest knowledge about what a roller coaster ride is like come about? The person has to take a ride. Similarly, we have to provide learners with ***evaluative experience*** of the right type. The ride is more that just the experience of having; it opens the door to future discourse, judgment. It fills in different kinds of knowledge blanks.
- Need to bring students into the guild; to **look at entire works through evaluative eyes.**
    - Holistic versus criterion-by-criterion.
    - Ego-involving versus task involving; inferences about inputs (such as effort) are irrelevant to quality determination and should have no influence on it.
    - Judging quality versus design-and-construct. Different domains of expertise. Developing evaluative expertise is a necessary but not sufficient condition for being able to produce quality works consistently.
- Where does **feedback** sit in all of this?
    - It is an essential part, but it is never enough. ***This is the case regardless of the quality or promptness of feedback***.
    - Design of teaching and learning environments must provide for **all** necessary elements.
    - This implies **standing back some from the traditional feedback model**, and distributing energy and resources across the requirements of description, exemplars and shared experience, with a view to developing sophisticated tacit knowledge.
    - To do so raises not only a structural but also a cultural challenge.
- Fact: Teachers typically make hundreds of qualitative judgments routinely each year as a normal part of their teaching responsibilities. They bring the accumulation of judgments about broadly similar responses in the past, and apply this knowledge to new student works. This activity exposes teachers to a wide variety of ways in which the students argue, describe, compare, evaluate, create, analyse, synthesize and solve problems.
    - Multiple appraisals give access to students' imaginations and strategies, providing vicarious experience about the challenges of production.

- Explanations for teachers' judgments invariably make use of ***criteria*** – which are constitutive elements of evaluative discourses – routinely invoking whichever criteria are salient to a particular judgment.

- Extensive exposure to multiple student works gives rise to two distinct types of knowledge.
  - Knowledge of the ***range*** of overall quality of the entire set of works (extended by the range of works the teacher has previously appraised). This exposure plays a significant role in the formation and maintenance of the teacher's abstract concept of quality, which carries *across* assessment events. Students often do not bet this exposure.
  - Knowledge of ***comparability***, framed by each subset of student works that are of about the same quality but are different from one another. Any given level of quality has many potential 'expressions'.
- Quality and comparability can be represented as two dimensions or axes, which define a 2-D space within which each student work can be located at a unique position. In marking a batch of student works, the assessor progressively populates that space with judgments about real cases.
- This is an exceedingly rich environment; students normally have limited or no experience of anything similar. How could we arrange for students to experience such an environment for themselves, so that feedback is not, unaided, expected to carry a load (as a key instrument for improvement) which is impossible for it to bear?
  - We need to 'bring students into the guild' of people who can also inhabit this 2-D appraisal space. Otherwise they are consigned to remaining mere consumers of feedback.
  - How can we provide learners with appraisal experience that is broadly similar in scope and kind to the teacher's?
- Main tool is ***peer assessment***…
  - … but not peer assessment as ***routine activity or busyness***. It has to be purposeful peer assessment that is designed with a ***clear pedagogical intent***.
  - Specifically, three classes of concepts set the metalearning agenda: response genre, quality, and criteria. These need to be explicitly addressed through structured peer assessment. (The order in which they are listed here is hierarchical, not accidental.)

Response genre as a concept.

- 'Response genre' refers both to the type of response set out in the assessment task specifications within which the teacher expects the students to respond, and to the type of response actually submitted by a student.
- Response genre is an issue because many students submit works that do not match the genre specified in the assessment task, and seem oblivious of this fact.
- Examples of response genres include solution, proof, algorithm, demonstration, extrapolation, and scenario.
- Giving credit for a work that does not match the required genre potentially has three negative side effects: the intent of the assessment task is subverted; the student's action is rewarded and reinforced without requiring a change in thinking and practice; and the student gains credit through avoiding a worthwhile educational outcome.

- Not all teachers intend assessment task specifications to be taken literally, and many use familiar forms of words out of habit. This sometimes occurs when academics focus primarily on the *content* they wish to test and pay only peripheral attention to *what students should do with* that content. This can carry through to marking: work that shows comprehensive coverage may be rewarded highly, regardless of its genre.
- In many fields, high-level academic learning can be demonstrated by the learner's ability to tackle diverse rather than routinised intellectual and practical problems. Assessing this requires that learners be able to assemble different selections of, and angles on, the substantive

content in ways that manifest themselves in a variety of response genres. Explicitly using the term *response genre* with students can provide the pragmatic basis for signalling it as a pre-emptive requirement. Unless a work is constructed within the designated genre, there is no answer to the question of how well it addresses the assessment task as specified.

- Students who appraise peer responses through the lens of response genre typically discover for themselves that many of those responses do not strictly address the set task. Conforming to the nominated response genre is seen not only as being required by the task specifications but also as defensible.

Quality as a concept.

- 'Quality' refers to the degree to which a work comes together as a whole to achieve its intended purpose. Determinations of quality look beyond superficial differences in the forms of individual works and into the deeper, subtler and more abstract aspects.
- This concept applies in disciplinary and professional contexts where complex learner productions are non-standardised, and expected to be so.
- The concept of quality is abstract and typically difficult for students to grasp. When complex phenomena are being evaluated, quality is often determined configurally rather than as the 'sum' of 'measures' of its components. Such holistic 'all-things-considered' judgments may amount to more, or sometimes less, than results 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 describe and account for in the particular.
- Not uncommonly, quality is difficult or impossible to reduce to explicit or declarative form, that is, to describe or specify wholly in words.
- In judging quality, considerable importance should be attached to a *holistic* judgment about it – attending to multiple criteria simultaneously in a configural way for the work as a whole.
- Quality often cannot be adequately promulgated in advance using, say, rubrics and criteria sheets. Students need to be exposed to, and gain experience in making judgments about, a variety of works of different quality, to provide explanations for those judgments, and to give verbal feedback on how the work could have been improved. As indicated above, they need to populate the 2-D appraisal space, as does the teacher. This points to the student need for planned rather than random exposure to exemplars, for experience in making judgments about quality, for verbalised rationales, for accounts of how a work could have been done better, and for conversations among teachers and students.
- Together, these provide the means for students to form a roughly similar concept of quality as the teacher possesses, and in particular to understand what makes for high quality. Although providing these experiences for students may appear to add another layer to the task of teaching, it is possible to organise peer assessment as a direct approach to pedagogy (Sadler, 2009a).

Criteria as concepts.

- A 'criterion' refers to a property or characteristic that is useful in the context of appraisal (of quality).
- Many criteria are abstract; this makes them problematic for students until they become competent users of them, which often requires some sophistication or fineness in judgment. These abstract criteria are concepts that do not have sharp boundaries, so they have to become known in the same ways in which concepts are formed by individuals and then shared.
- Criteria need to become part of the student's appraisal vocabulary, to enable them to explain (justify) their judgments, and also to reason through qualitative judgments in their own minds.
- Students need to understand what the criteria 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 in a particular appraisal.

- Telling can inform and edify only when the meanings and implications of the terms and structure of the communication are understood by the student as message recipient.

Tacit knowledge

- 'Tacit knowing' (Polanyi, 1962) consists of the subtle and often elusive aspects of knowing that people develop and share primarily through engaging in common experiences. It is called tacit because it is difficult or impossible to articulate adequately, that is, to express in explicit or propositional verbal form. Connoisseurship typically relies heavily on it.
- A learning environment that attends to developing students' conceptual understandings of response genre, quality and criteria will, as a consequence, extend the students' tacit knowledge. It cannot be 'taught' directly.
- Making numerous and purposeful peer assessments is the crucible within which these three concepts can be seen to react and interact. This activity gives rise not just to an appraisal 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.
- 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.

Conclusion

- If we want our assessment to be truly and professionally formative for students, even meticulously constructed feedback will not deliver. We have to change the way we think and act.
- Taking up such a challenge is an exciting venture, but produces surprising rewards. And it need not be more labour intensive. Teach less, teach differently. Students learn more in new and deeper ways.

References

Polanyi, M. (1962). *Personal knowledge: Towards a post-critical philosophy*. London: Routledge & Kegan Paul.

OTHERS

Sadler, D. R. (1983). Evaluation and the improvement of academic learning. *Journal of Higher Education*, **54**, 60-79.

Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice,* **5**, 77-84.

Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment and Evaluation in Higher Education*, **30**, 175-194.

Sadler, D. R. (2009a). Indeterminacy in the use of preset criteria for assessment and grading in higher education. *Assessment and Evaluation in Higher Education*,

Sadler, D. R. (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.

Sadler, D. R. (Draft manuscript). Beyond feedback: Developing student capability in complex appraisal. (Available on request.)