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, May 2009.

Outline. *Assessment should lead to improved learning – but often fails to. The reasons are many, and obviously include both teacher-based factors (the design and execution of assessment strategies) and learner-based factors (student dispositions or lack of engagement). Our focus will be on what we as academics can do to make assessment work more effectively for students. We also grade student achievement (summatively). Sometimes both formative and summative purposes and practices become confused. How can all this be unscrambled so that assessment works well for academic teacher-assessors and for the student-learners?*

Part I

Terms

- *Assessment*: any appraisal (or judgment, or evaluation) of a student's work or performance. (Older term: *evaluation*).
- *Formative assessment*: Using information derived from student responses to assessment tasks in order to shape, improve student achievement. Responses may be called performances, pieces, or works. Usually it is the teacher who makes the definitive judgments about the quality of student responses. Teaching generally, and formative assessment in particular, are intended to short-circuit the randomness and inefficiency of trial-and-error learning. Etymology & common usage: *formative* associated with forming or moulding something, usually to achieve a desired end; 'formative years' in childhood.
- *Summative*: summing up or summarizing the achievement status of a student at a given point in time; geared towards reporting at the end of a course especially for purposes of credit or certification. Essentially passive; summative normally has no immediate impact on learning, although some possible in long term.
- Conceptual distinction between *formative* and *summative* is due to (Scriven, 1967).
- Primary substantive distinction between formative and summative relates to **purpose and effect**, not to **timing**. Ordinarily, cannot identify the intent from just the structure of assessment tasks.
- Many of the principles appropriate to summative assessment are not necessarily transferable to formative assessment; distinctive conceptualization and approach is required for formative.

Feedback

- *Feedback*: usually thought of as a (or *the*) key element in formative assessment – information about how successfully something has been or is being done. (Sadler, 1989)
- Few physical, intellectual or social skills can be acquired satisfactorily simply through being told about them. Common approach: practice in a supportive environment incorporating feedback loops. Usually includes a teacher who knows which skills are to be learned and who can recognize and describe a fine performance, demonstrate a fine performance, and indicate how a poor performance can be improved. Feedback can also be defined in terms of its *effect* rather than its informational content: 'Feedback is information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way' (Ramaprasad, 1983, p.4). This alternative definition emphasizes the system-control function.
- Elements of feedback loops: sensor (receptor), comparator, effector.

- Distinction between *positive* & *negative* feedback loops (Wiener, 1948). Background: computer science & neurological science (MIT & Mexico).
- ***Negative feedback***: homeostasis (Gk: 'same state'); system 'in control' through constant monitoring and correction as necessary; events or conditions too away from the reference level (or outside a bracket of levels) are corrected. Contrast: calibrated instruments for measurements.
- ***Positive feedback***: enhanced output; may be *escalation* (runaway; explosion; system collapse; out of control; Ponzi schemes; freeway development!) or *oscillation*.
- NOT called 'positive' <u>because</u> it is 'feel-good' (although this is the way learning theorists tend to use the term; also called 'ego-involving'.)
- EBay feedback: Positive in both feel-good and system control senses.
- Lots of systems feature feedback in one form or another: Engineering, physiological systems, neurological systems, organisational management, socio-political systems (democracies), environmental science (ecological systems; interdependent dependent organism populations).

Learning in higher education

- What we want:
  - BOTH Negative feedback.

    Unsatisfactory aspects are modified or rectified; problems identified are fixed; performance brought closer to the reference level; improved aspects associated with success or high quality are recognized and reinforced;
  - AND Positive feedback.

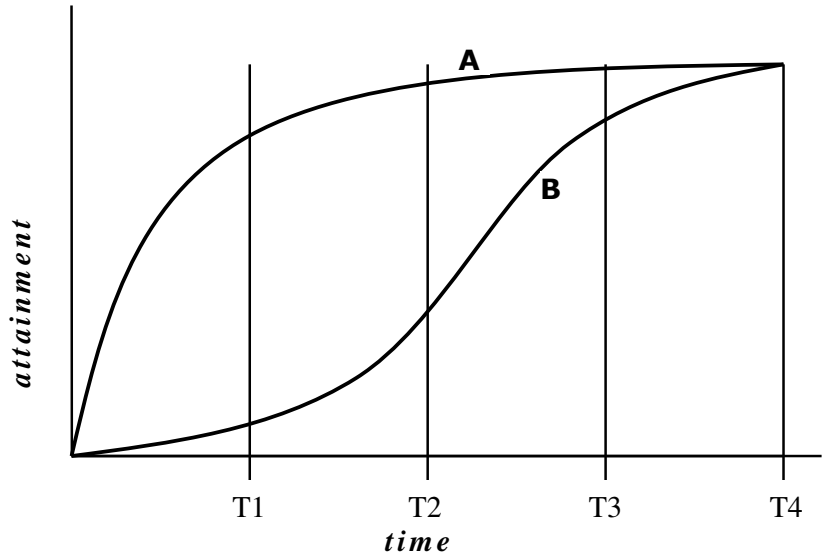
    Moving forward from present state of learning – academic growth; not a static system 'under control' but systematic learning; 'virtuous circle' thinking.
- Ramaprasad's definition is that information about the gap between actual and reference levels is considered as feedback *only when it is used to alter the gap*. If the information is simply recorded, passed to a third party lacking either the knowledge or the power to change the outcome, or is too deeply coded (for example, as a summary grade given by the teacher) to lead to appropriate action, the control loop cannot be closed and 'dangling data' substitute for effective feedback. When a grade or score assigned by a teacher acts as a one-way cipher for students, attention is diverted away from fundamental judgments and the criteria for making them.
- A bare grade may be counterproductive for formative purposes. Students need more than summary grades if they are to develop expertise intelligently. The indispensable conditions for improvement are that the *student* comes to hold a concept of quality roughly similar to that held by the teacher, is able to monitor continuously the quality of what is being produced *during the act of production itself*, and has a repertoire of alternative moves or strategies from which to draw at any given point. In other words, students have to be able to judge the quality of what they are producing and be able to regulate what they are doing during the doing of it.
- Summarizing that: for success, the learner has to
  - Possess a concept of the *standard* (or goal, or reference level) being aimed for;
  - Be *able to compare* the actual (or current) level of performance with the standard; and
  - *Engage in appropriate action* which leads to some closure of the gap.
- Shenstone: 'Every good poet includes a critick; the reverse will not hold' (1768, p.172).
- Formative and summative both at once?
  - At the technical level, YES. It should be possible to put the same data to different uses.
  - At the practical and human level (students' perceptions & reactions), NO! Summative subverts the formative function. Two reasons: (a) It's too late to do anything to retrieve that

particular situation; and (b) The next task to which the specific feedback could apply may be (i) different in type, and (ii) a full semester away.

- Problem: in many HE contexts, academics condition students to think that everything MUST count, or students won't take it seriously. Also, we may think students *deserve* it. Effort (or just activity) in → marks out. Bank the marks; withdraw account completely at semester end, see what the grade is. Students now expect this; we meet that expectation. System is strongly self-reinforcing.
- By definition, summative represents high stakes for grading. This significantly reduces the stakes for learning (the horse has bolted).
- Formative needs to be ***High stakes for learning***; ***Zero stakes for grading***.

Cumulative assessment of learning (Sadler, In Press)

- **In relation to course objectives**.
  - The usual way of expressing objectives is to prefix each set by a stem similar to this: 'At (or by) ***the end of the course***, students are expected to …'. Except for courses composed of independent sub-courses, cumulative assessment undermines that intention.
- **In relation to shape of attainment path**.
  - Attainment paths differ from student to student.



Two hypothetical attainment paths

  - Consider two hypothetical students; same course, same high level of knowledge and skill by the end of course. Student A grasps the material rapidly. Student B exhibits initial struggle with the early material but with rapid acceleration of knowledge and skills from about half way through.
  - Suppose there are four progressive assessments, equally spaced across the course, at T1, T2, T3 and T4. In diagram shown, weighting progressive scores equally and adding them makes B's aggregate about 65% of A's, despite identical achievement levels at course end. B's initial difficulties create permanent score deficits that cannot be offset or overridden by later catch-ups.

- The aggregate scores for these two students reflect both *final level of attainment* and their respective *attainment paths*. Changing the weightings changes the relative contributions of the attainment path shapes, *but not the principle*.
- Other attainment paths (not sketched): Student C has attainment path which is essentially flat and at a low level for most of the learning period, and then shows almost vertical rise between last two assessment points. Abrupt insights help everything fall into place ('Ah Ha' experiences, Simon, 1983). Student D has considerable background knowledge and experience on entering course – but does not grow much during it. Cumulative scores can exceed those of other students who perform at higher levels by the end of the course.
- Cumulative assessment raises practical, ethical and interpretive problems that flow from the decision to accumulate.

Summative assessment design

- Redesign assessment programs in courses to focus on actual level of academic achievement attained by the end of the course.
- ***Product focus*** is critical (personal, professional capital). ***Process focus*** is important for evaluating teaching techniques, and may be otherwise interesting as well – but is irrelevant to grading.
- The focus on the attained level academic achievement implies including no contaminants, such as effort or participation; the course grade must represent academic achievement status, and nothing extraneous.
- Think creatively about design and administration parameters so that the summative assessment program fits (i) the summative purpose, and (ii) the context.
- In particular, we need to get answers to three questions:
  - How could we ***best and most efficiently obtain the evidence of achievement*** required?
  - How could we develop different modes of assessment that foster the development of, and demand, ***high-order intellectual and professional knowledge and skills***?
  - How can we think beyond traditional assessment (for example, end-of-semester time-limited formal examinations) and still met the requirements for assessing each student's own work, where appropriate?

References

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28, 4-13.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, **18**, 119-144.

Sadler, D. R. (In Press). Fidelity as a precondition for integrity in grading academic achievement. *Assessment and Evaluation in Higher Education* (Available from author as manuscript)

Scriven, M. (1967). The methodology of evaluation. In Ralph W. Tyler, Robert M. Gagné, and Michael Scriven (eds.), *Perspectives of curriculum evaluation*: Vol. 1. AERA Monograph Series on Curriculum Evaluation (pp. 39-83). Chicago: Rand McNally.

Shenstone, W. (1768). On writing and books, LXXIX. In *Works: In verse and prose* Vol. 2, (3rd ed.). London: Dodsley.

Simon, Herbert. (1983). Alternative visions of rationality. In *Reason in human affairs,* ed. Herbert Simon, 3-35. Oxford: Basil Blackwell. (Series: Harry Camp Lectures at Stanford University, 1982). Edited and republished in *Rationality in action:* contemporary approaches, 1990, ed. P. K. Moser, 189-204. Cambridge: Cambridge University Press.

Wiener, N. (1948). *Cybernetics or control and communication in the animal and the machine*. Paris: Hermann et Cie; Cambridge, MA: MIT Press.

OTHERS

Sadler, D. R. (In Press). Grade integrity and the representation of academic achievement. *Studies in Higher Education*.