Self- and peer assessment

Amy Hamilton and David D Curtis
School of Education
Faculty of Education, Humanities and Law
Teaching and Learning Forum
3 December 2012

Assessment: A process for eliciting evidence

Assessment is:
- … a process for obtaining information that is used for making decisions about students; curricula, programs and schools; and educational policy (Nitko & Brookhart, 2011, p.3)
- … a process in which educators use student responses to tasks to draw inferences about students’ knowledge and skills (Pellegrino, Chudowsky, & Glaser, 2001, p.20)

How do self- and peer assessment fit into pedagogical and assessment practices?

We distinguish between formative and summative assessment.

Formative assessment

Informs learning
- initial or diagnostic assessment
- carried out throughout a course or project
- provides information on the progress of learning, and therefore informs both learner and teacher about what learning and teaching actions are required
- might be done by a teacher or peer or the learner
- not used for grading purposes

Crucially, if formative assessment is to provide information that can inform learning, it must involve feedback that focuses on the gap between current and desired levels of performance and specifically on the actions (by learner and teacher) that are required to close that gap.

Summative assessment

It is essential in education as we are required to certify that graduates meet prescribed standards to enter professional practice. It:
- is used to pass or fail a student
- is used to grade or rank a student
- occurs at end of learning unit
- is used to compare student performance against standards
- is used to tell students/parents/administrators/employers what students have achieved (what they know and can do).

Effective assessment

Whether formative or summative, assessment must:
- Have clear goals and purposes
- Be challenging and achievable
- Have clear and comprehensible success criteria
Success criteria must be standards-referenced, but students are unlikely to be able to interpret standards, which are necessarily abstract, and apply them to immediate tasks. Therefore, academic teachers must provide the translation from standards to criteria. (We can hope that, over time, students will begin to perceive the relationship between standards and criteria and be able to perform the translation themselves) (Hattie, 2003, 2005; Hattie & Timperley, 2007; Ramprasad, 1983; Sadler, 1989, 2005).

Effective formative assessment requires feedback. Effective feedback is:
• task-focused (not person-focused)
• timely – given very shortly after the assessment task
• relates the response to the criteria and performance standards
• identifies explicitly any gap between the response and the standard.

Feed-forward is a form of feedback that comments on task performance and provides advice about future performance (Hattie & Timperley, 2007).

**Relationship between formative and summative assessment**

Some authors claim that individual assessment tasks can serve both formative and summative purposes (Black & Wiliam, 1998). However, there is contrary evidence. Feedback as grades trumps all other feedback – if grades or marks are given, students do not attend to any other feedback (Black & Wiliam, 1998).

![Figure 1 Possible relationships between aspects of formative assessment and summative assessment](image)

**Warning: formative assessment is lethal – it could kill you**

If you want to do the right thing and provide extensive and detailed feedback in order to promote learning, the workload involved will kill you.

A possible solution exists. Biggs (1999) makes the important point that learning occurs because of the conscious and purposeful cognitive actions that students undertake. If we apply this to formative assessment and encourage students to examine criteria and to make judgments about their work (their own and peers’), we may be doing them a greater favour than if we do all the work in making those evaluations and describing and justifying them by way of feedback.

**What is the rhetoric?**

Assessment (summative) is something that is done to students. It is accepted by students who are successful, but is disempowering for less successful ones. Their problem is that
assessment is a closed book and they are excluded from the community of practice (Boud, 2000, 2002).

Part of a solution to this involves inviting students into the community of assessment practice and encouraging them to understand and apply assessment standards and criteria. While there are advocates of summative self-assessment, it places students in conflicted positions and is seems they are unlikely to be able to focus on standards and criteria and to make objective judgments.

**Self-assessment**

Many authors have proposed self-assessment as a way to escape the reliance on judgments by others, to engage students in understanding and internalising standards and success criteria, and to empower students to make judgments about their own performance for themselves (Biggs & Moore, 1993; Boud, 1995, 2002).

Ward, Gruppen and Regehr (2002, p.63) begin their argument for self-assessment in medicine with the following:

As with any profession that operates under the principles of self-regulation and autonomy, the competent physician must be a self-directed, lifelong learner (Moore and Cordes, 1992).

In any professional (or vocational) education program and probably others, we need to ensure graduates have the capacity for accurate reflection and learning, because that is one of the ways they will sustain and enhance their professional competence following graduation.

The evidence for the effectiveness of self-assessment is strong (Hattie, 2009, and see below).

However, self-assessment is incompatible with summative assessment and there is evidence to suggest that self-assessments are weakly to moderately correlated with the judgments of more-expert others (Lew, Alwis, & Schmidt, 2010; Ward, et al., 2002).

**Peer-assessment and collaborative learning**

Peer assessment can be used for formative or summative purposes. Most reports refer to its formative value.

We are not sure about the relationship between collaborative learning and peer-assessment. It seems logical that peer-assessment will be accepted most readily by those students who are comfortable in collaborative learning contexts, although that is probably no guarantee that students will accept or listen to feedback from peers. It seems likely that many individual characteristics, including prior achievement, will influence acceptance of peer feedback.

We are not sure about the mechanisms by which peer assessment works. Topping (2009) suggests that the greater immediacy of feedback by peers compared with delayed feedback from teachers might be a contributing factor. Falchikov (2001) distinguishes peer assessment from peer feedback; in the former involving marks or grades are assigned while in the latter, only comments are provided. Wiliam (2011) presents evidence from teacher assessment that the provision of grades undermines the positive influence of any feedback. If this applies to peer assessment, it seems that peers should not be asked to assign grades but to focus on providing comments.

**What is the research evidence?**

Hattie (2009) reviews over 800 meta-analyses of a vast number of studies of educational interventions and examines their effect sizes – a standardised way of reporting the instructional effectiveness of an intervention. Effect size answers the question ‘How big a difference does the intervention make?’
Effect sizes\(^1\) are reported on a scale that can be negative (the intervention makes the situation worse), through zero (it makes no difference) to figures potentially greater than unity. The following effect sizes and their qualitative magnitudes are:

- Very small to small: 0.1 to 0.3
- Medium: 0.3 to 0.5
- High: 0.5 to 0.8
- Very high: >0.8

He reports an average effect size for formative assessment of 0.9, in the very high range. If we want to improve teaching effectiveness and enhance student learning, we do need to embrace formative assessment.

Some other interventions associated with formative assessment and their effect sizes are:

- Reciprocal teaching: 0.74
- Feedback: 0.73
- Metacognitive strategy use: 0.69
- Cooperative learning: 0.59
- Worked examples: 0.57
- Concept mapping: 0.57
- Goal setting: 0.56
- Peer tutoring: 0.55
- Self-assessment: 0.47
- Time on task: 0.38

The last of these, time on task, is shown as a reference value. It has been held up as a gold standard in educational interventions. It has been the subject of classroom research since at least the 1960s and is held up as an objective in classroom practice: Do your classroom practices and instructional actions lead to increased student time on task?

We have shown the effect sizes for a range of interventions that are related to formative assessment and self- and peer-assessment. They are all greater than the time-on-task effect.

**Where there is a lack of evidence – research opportunities**

Although Hattie (2009) reviews extensive evidence, our understanding is far from complete. We do not know how effective the various methods are in the disciplines we teach.

There are opportunities for Discipline Based Education Research (DBER) within each of our disciplines. This research takes on board findings from other disciplines and integrates them with the norms (the priorities, worldview, knowledge and practices) of the discipline. DBER recognises what expertise looks like within the discipline, understands that students are novices, and creates learning pathways towards expertise. Assessment is critical in those pathways. Summative assessment is about certifying that graduates have attained a sufficient level of professional knowledge and skill to enable them to begin professional practice. Formative assessment, in its various manifestations, can be used to illuminate those learning pathways.

We do not know much about important relationships, e.g. between peer assessment and collaborative learning or the groups of students who may gain the most from peer-assessment and collaborative learning.

Some evidence is emerging that in peer assessment, the student assessors gain more from the experience than those who are assessed (Topping, 2009). It is possible that this outcome arises because, in assessing the work of another student, the student assessor has the

---

\(^1\) Hattie uses Cohen’s ‘d’ as an effect size measure. There are about 10 different measures of effect size.
opportunity to understand and apply assessment criteria to real pieces of student work and that they do this without the interest conflict that arises in self-assessment.

There is speculation and some evidence that peer assessment is compromised by existing peer relationships in class. Topping (2009, p. 24) wrote:

Social processes can influence and contaminate the reliability and validity of peer assessments. Peer assessments can be partly determined by friendship bonds, enmity, or other power processes, the popularity of individuals, perception of criticism as socially uncomfortable, or even collusion to submit average scores, leading to lack of differentiation.

We need to identify peer assessment practices that avoid the threats of counter-productive social processes while harnessing those that are implicated in collaborative learning.
References


