

Can we raise the level of debate on teacher assessment?

Rationale

The potential of summative assessment by teachers is presently being taken seriously by education policy makers. However, the level of debate is still too low to support effective decision-making. The ASF group can make an important contribution by helping to raise the level of debate.

First, there is a tendency to treat *all* teacher assessment as *good* assessment, and not to draw important distinctions between vastly different teacher assessment models. For example, consider the following recommendations from the Working Group on 14-19 reform (see Annex 1 for a more thorough analysis):

A model involving radical change?	A model involving little change?
"... teachers' assessment of the ongoing work of the course (not formal coursework specified by an awarding body)..." (151)	"... a system of assessment led by teachers, tutors and trainers exercising their professional judgement. For many this would not be a new role." (150)
"... overall judgement of each young person's progress is continually updated..." (152)	"... every teacher is already engaged in setting and marking assignments, tests and practical projects..." (150)
"... observations..." (12, also 152)	"The system of professional judgement... would be characterised by... time-limited tests and examinations... and practical and written tests" (152, also Rec. 20)
"... project and portfolio work..." (141, also Rec. 20)	

Second, there is a tendency to fail to prioritise between:

- different purposes for assessment (e.g., certification, accountability)
- different intended impacts of assessment (e.g., better feedback to managers, better professional development of teachers)

Obviously, it is not possible to design an optimal system unless we have a clear view of what it needs to achieve. This means that we have to prioritise between different purposes and impacts as explicitly as possible. (See Annex 2 for a more thorough analysis of a range of motivations for assessment reform as presented in the report of the Working Group – note that these seem not to be prioritised other than, perhaps, the need to reduce the ill-defined 'assessment burden'.)

The ASF Group might help raise the level of debate by examining in more detail:

- how we think about teacher assessment
- what we produce for policy makers
- how we sell our message to policy makers

I think it would be great for us to be in the position of presenting very clear messages to policy makers along the lines of: if you decide on that model/system rather than

Draft for discussion – restricted circulation.

this one, then this is likely to go right and that is likely to go wrong. Then at least we can say ‘I told you so!’

Theoretical clarity – how we think about teacher assessment

I think that certain teacher assessment models may be *bad* for many impacts and purposes (e.g., frequent unstructured quizzes?) while other teacher assessment models may be *good* for many impacts and purposes (e.g., continuously updated structured portfolios?)... does the Group agree? ... can we identify good and bad models – either generically or through specific example? ... or would identifying ‘generally’ good or bad models detract from the real task of finding the right model for the desired impacts and purposes?

I think that we need explicitly to prioritise different assessment purposes and impacts before making claims about whether particular assessment models will deliver results of sufficient technical accuracy... does the Group agree?... do we agree which purposes and impacts ought to be central and which ought to be peripheral (how thoroughly have we thought it through)? ... do we know how different policy makers would respond (how thoroughly have they thought it through)?

I think that it’s extremely hard, if not impossible, to determine the defensibility of an assessment system unless we consider both the assessment model and the context within which it is situated... does the Group agree? ... should we be debating assessment systems rather than assessment models? ... are systems based on hybrid models ever satisfactory?

Methodological clarity – what we produce for policy makers

Raising the level of debate is going to be very hard, because good debate – in this domain – is highly complex. How can we model good debate for policy makers?

Can we provide simplified formats for illustrating the more complex aspects of the debate (or will we risk over-simplifying it)? For example, ‘defensibility grids’ along the following lines (this example is purely for illustration and is not intended to be definitive in any sense):

- primary assessment purpose = school-leaving certification
- results will not be used for accountability/monitoring purposes

- primary educational impact = supporting the development of formative assessment practices

		narrow assessment goal	
		– ve	+ ve
broader educational goals of assessment reform	– ve	short/narrow test	test + tick-list TA
	+ ve	unstructured holistic TA	structured portfolio TA

Draft for discussion – restricted circulation.

Alternatively, could we draw up grids which summarise what we are likely to win or lose under different assessment models, in terms of purposes and impacts? Or, similarly, highlight the key threats to intended purposes/impacts under each model/system, alongside tips for minimising them? Has this been done before?!

Rhetorical clarity – how we sell our message to policy makers

If we were to begin to elevate certain teacher assessment models/systems above others, might that disunite the 'movement'? Would that be a bad thing?

Should we present policy makers with systems which fit *their* preferred purposes and impacts, or with systems which fit *our own*, or both? (Again, even within the group, do we actually share the same values?)

Should we argue our case by selling the virtues of summative teacher assessment, or by foregrounding the evils of tests and exams? In particular, should we preface all of our arguments by highlighting the technical inadequacy of tests and exams? My view is that this is the wrong way around. I think that any argument has to be mounted in the following direction:

1. specification of intended positive educational impacts
2. negotiation of intended assessment purposes
3. identification of appropriate model for achieving impacts and supporting purposes

Of course, this might result in recommending a sole reliance upon tests/exams, given certain intended impacts and purposes! So be it. Given a different set of intended impacts and purposes, though, a sole reliance upon teacher assessment might be recommended.

Why this way around? First, I think we need to win hearts (by selling the positive impacts of teacher assessment) not minds (by preaching the limitations of tests/exams). Second, by preaching certain limitations of tests/exams (e.g., reliability/validity) we implicitly elevate certain purposes/impacts which may be less valued by others (e.g., the accountability purpose despises comparability error more than reliability or validity error). Third, there's no guarantee that teacher assessment will demonstrate greater technical accuracy than tests/exams (and you'd never get definitive evidence even if it was). Where teacher assessment is (presumably) going to win definitively is in terms of impacts; so shouldn't we focus on what we clearly stand to gain by embracing teacher assessment (or what we clearly stand to lose by not embracing it)?

Finally, should we focus our criticisms upon the underlying purposes/impacts or (again) upon specific assessment models? This is similar to my earlier question... are we against the stakes or the tests? I'm against the stakes rather than the tests... and I don't expect policy makers to embrace the full potential of teacher assessment until the present performance table accountability framework is no more.

Paul Newton
Research and Statistics Team, QCA
22 December 2004