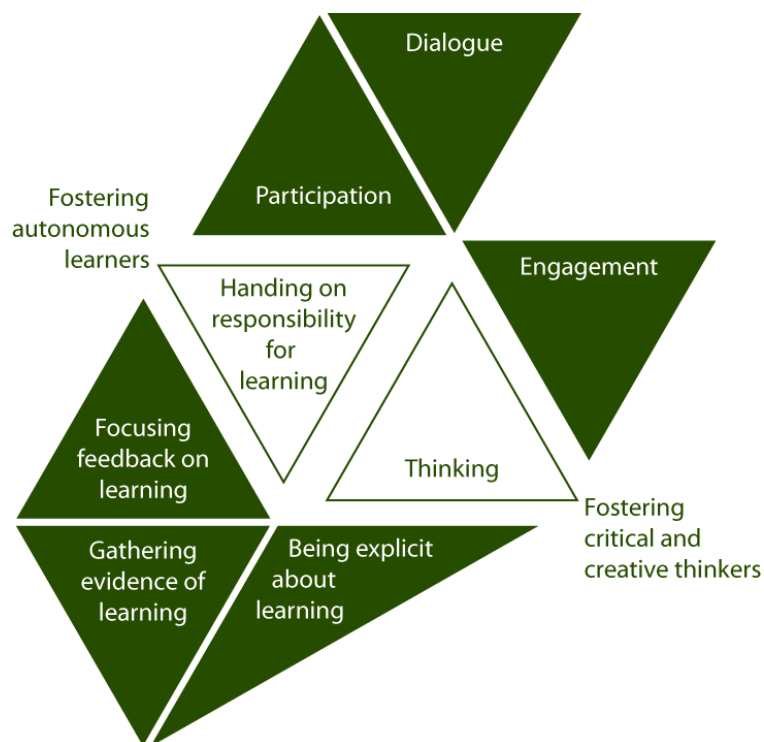


Embedding A Curriculum for Excellence in the Classroom

*A Highland CPD Journey
2004 - 2009*



The Highland Council
Learning and Teaching CPD Reflection Framework

Professor Paul Black

*An analysis of the issues raised by the Highland Council's recent papers on its
policies for assessment and pedagogy*

Reflective Professionals and Thinking Children

Preface

The Highland Council CPD Reflection Framework and The Highland Model

Over the last five years, groups of teachers and schools in The Highland Council have been on a CPD journey exploring ways of fostering the CfE capacities and raising achievement. Its aims were to:

- develop a coherent conception of formative assessment
- explore the links between formative assessment and approaches to making thinking explicit as a powerful way of fostering the CfE capacities
- encourage teacher learning communities which foster reflective professionals through a variety of approaches to CPD
- raise achievement, motivation and confidence

The journey began in 2004 by exploring the links between formative assessment and thinking skills. A CPD Reflection Framework was then developed to explore formative assessment as a coherent, practical way of engaging with the values, purposes and principles of Curriculum for Excellence. Groups of Highland practitioners and SMT began to work with Eric Young to explore what it means to embed and extend formative assessment, with Mark Priestley on managing and sustaining change and with Professor Robert Fisher on linking formative assessment and thinking skills.

The Highland Council CPD Reflection Framework was designed to support the emerging CfE policy documents. The Highland Model was built around a clear set of principles - *participation, dialogue, engagement and thinking* - which became the basis for developing a distinctive model of effective learning in the context of Curriculum for Excellence in which the principles and practice of formative assessment are used to help pupils take greater responsibility for their own learning.

The CPD Reflection Framework www.hvlc.org.uk/ace/aifl/ has been extended to include additional resources including case studies and reflection booklets and a new unit (Unit C3) explores how The Highland Model can be used to help teachers meaningfully engage with CfE learning outcomes.

The Highland FLaT Project and SQA Evaluations

The model also informed another stage of the journey, the Highland Council Future Learning and Teaching (FLaT) project, *Embedding a curriculum for excellence in the classroom (06-08')*. The project involved teachers from three clusters working collaboratively to explore ways of fostering the CfE capacities using the principles and practice of formative assessment. As the journey progressed, five subject ASGs were set up to investigate how pupil peer and self assessment could be used to address the formative / summative tension in the upper secondary. The SQA commissioned a review of this aspect of the Highland Journey. Both the SQA and FLAT evaluations have been very positive.

Three engagement papers

As part of the formative evaluation of the Highland journey it was agreed that three engagement papers would be produced to stimulate discussion regarding possible next steps. The first paper explores how the pedagogical model was co-created by groups of teachers collaborating with a number of researchers. The second paper explores ways in which teacher learning communities can be sustained. The third paper *Policy, Research and Practice* explores some possible implications for the policy and research communities through the prism of 4 contemporary tensions related to the role of assessment and thinking which were a central feature of the Highland journey (i) principles vs. strategies (ii) the formative / summative tension (iii) intelligent accountabilities and (iv) learning how to learn and subject knowledge.

Following the publication of these evaluations, Professor Paul Black, co-author of *Inside the Black Box*, agreed to identify areas for discussion to inform the development of learning, teaching and assessment policies in Highland Council. His paper draws on the above evaluations and engagement papers; in particular, it addresses the 4 tensions from the third engagement paper.

Further information on the Highland journey can be found at www.hvlc.org.uk/ace/aifl/ or from kevin.logan@highland.gov.uk

An analysis of issues raised by the Highland Council's recent papers on its policies for assessment and pedagogy

Professor Paul Black, Kings College London

1 Introduction and overview

This paper aims to set out my responses to several of the issues raised by the Highland Council's three papers which are concerned, respectively, with Building Teacher Learning Communities, with Sustaining Teacher Learning Communities, and with Policy, Research and Practice, and with the Council's CPD reflection framework which discusses the development of a pedagogical model. In my Section 2, I explore some of the problems about formative assessment and classroom learning which can be seen in these papers, mainly in relation to the pedagogical model. Then in Section 3, I discuss the problems for policy and practice raised by the various purposes served, at different levels in the system, by summative assessment. In section 4, I address more directly the issues raised by the need to support teachers in their involvement in and implementation of changes in policy and in practices. Section 5 presents a general plea for coherence across the policy making system, whilst a closing section summarises some of my conclusions.

2 Formative assessment, learning and pedagogy

2.1 Many authors have offered definitions of formative assessment, and clarity about its meaning is clearly essential. The definition to be used here is taken from Black and Wiliam (2009)¹:

Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited.

The term 'instruction' as used here includes what would be implied by pedagogy but is broader to ensure inclusion of learners (pedagogy undertaken by pupils would sound strange). The focus on decisions is to emphasise the central aim of the information. Any decision may lead to taking action now, or in the future, or not at all - each could, depending on context, be a decision helpful to the learning. Such decisions will usually be made by the teacher, but the definition also include peers, or the individual learner, as agents in making such decisions, and indeed one of the aims might well be to tip the balance so that the teacher hands-over responsibility for decisions to the learners.

2.2 Another aspect of clarity in discussing formative assessment is that it was originally described in terms of a collection of practices. As understanding of its implications has grown, a deeper and more unified view has developed. This may be explained by the

¹ Black, P. & Wiliam, D. (2009) Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, **21**(1), 5-31.

following list given by Black and Wiliam (2009) in which they formulated the main strategies as:

1. clarifying and sharing learning intentions and criteria for success;
2. engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding;
3. providing feedback that moves learners forward;
4. activating students as instructional resources for one another; and
5. activating students as the owners of their own learning.

The five types of activity, which were identified earlier as the starting point for work on work on formative assessment, can now be seen as means of enacting these key five strategies. Whilst all of them could contribute to the first, classroom questioning is the main way of implementing the second, and comment-only marking is a particular way that teachers might achieve the third. Similarly, peer- and self-assessment are activities that might be used to pursue the fourth and fifth respectively: these are particularly relevant to the development of students' own capacity to learn how to learn and to learner autonomy (Black et al. 2006)².

The last of the original types of activity, the formative use of summative tests, can address both the first, fourth and fifth, but it has more complex implications which are discussed further in sections 2.6 and 3 below.

2.3 Pedagogy: Because work on formative assessment was promoted and developed without a clear view of its role in a comprehensive theory of teaching and learning, its interdependence on other aspects of pedagogy has not been spelt out. A teaching strategy should start with a clear *purpose*, then frame *classroom-based tasks* to work towards those purposes, then so *conduct the interactions* in and around the implementation of those tasks that pupils develop understanding and confidence, and include some foresight about how to *review progress* from time to time to determine how the strategy may have to be revised. Formative assessment is concerned with the *interactions*. It can only be effective if the other components are developed in harmony.

2.4 Purpose: The choice of *purpose* has two aspects. This first is the task of introducing learners to their cultural heritage, a key function if society is to build on, and enjoy the fruits of, past achievements. This is often misrepresented, even denigrated, by use of the term 'content': young people have a right of access to great literature, to drama, to the insight into the natural world of science, to the interplay of argument about ultimate beliefs and morality, and so on. None of these is in opposition to the development of deep understanding - indeed they are worthless without such understanding. The second aspect is to develop powers of the mind, and of physical action, again often mis-represented as 'skills'. Critical awareness of one's own thinking, development of that awareness to achieve a deeper level of thought, together with co-ordination of thought with bodily action are all involved.

Broad sweeping aims can be attractive and unhelpful. An example is "learning to learn": who could be opposed to this purpose? But what does it mean? If I learn mathematics, does that

² Black, P., McCormick, R., James, M., & Pedder, D.. (2006), Learning how to learn and assessment for learning: a theoretical inquiry. *Research Papers in Education*. **21**(2) 119-132..

mean that I have learnt to learn mathematics, and if I have, does this mean I am now good at learning literary appreciation, or at learning how to run the 100-metre hurdles?³ More concrete purposes, grounded in the long traditions of the diverse academic disciplines, have the advantage of coherence of constructs, supported by explicit examples (e.g. you are good at mathematics if you can tackle these three examples of real-life problems), as well as being the components of the cultural heritage which society ought to pass on to successive generations.

In this broader view, the presumed conflict between ‘content’ and ‘skills’ seems naively misconceived. The problem is how best to manage the interaction between the two aspects. It has indeed been recognised that transfer of understanding and ways of learning across different contexts is difficult, not least because the change of context usually means that these qualities have to be modified or re-configured to meet the new context. Even in two subjects as close as mathematics and physics, the differences in the ways of thinking are as important as the similarities.

As pupils become more engaged in examining and raising questions about any issue or event, they will test to the limit the understanding of their teachers. Deep understanding of the material with which they are engaging their pupils is far more essential than it is for ‘delivery’ styles of teaching. So teachers have to be learners themselves, keeping abreast of their subjects and developing their “pedagogical content knowledge”. Collegial exchange and support can be very helpful here.

None of this is to argue that cross-subject activities are unimportant - indeed, many important and everyday problems can only be tackled effectively by drawing on the resources of the different subject disciplines, and the specific contribution of each subject can be better understood in a cross-subject project. Examples of how this can be done in imaginative ways have been explored in the past: they require the commitment of some of the timetable times of several subjects to a collaborative project.⁴

2.5 *Formulation and presentation of tasks*: The skill here is to formulate, and present to pupils, tasks (questions, reading, practical activities) which have the potential to engage and challenge them in such a way that their responses will provide critical evidence about their prior understanding and will open up a dialogue to develop that understanding. Often, it is only as they see the outcomes, in terms of formative engagement following or not following from their various tasks, that teachers can make progress in their own learning, including the insight needed to anticipate the potential for learning of the tasks they formulate or select.

³ This issue has been explored in depth by the TLRP Learning How to Learn project - see Black et al. 2006 quoted in footnote 2.

⁴ Murray et al. (1994) *Managing Learning and Assessment Across the Curriculum : a Guide for Secondary Schools*. H.M.S.O.: London. Paechter, C. (2000) *Changing School Subjects: Power, Gender and Curriculum*. London: Open University Press (chapter on Design and Technology as a new school subject).

2.6 *Interaction with and between learners*: This is where formative assessment is located: here, it is distinguished by the characteristic that the learning takes place in close co-ordination with the act of assessment. Its conduct requires careful steering of the dialogue by the teacher, with a skilful control which is sufficiently light and flexible to give pupils time and opportunity to think and interact, but sufficiently firm to keep the dialogue moving towards the intended purpose. Feedback in both oral and written dialogue, and involvement of pupils in small groups and in self- and peer-assessment are all essential here. Pupils will often produce surprising and puzzling responses, and the teacher has to decide whether and how to both take these seriously and to follow them up. This is a formidable problem for teachers, since [their](#) feedback needs to be constructed in the light of some insight into the mental life that lies behind the student's utterances. As one researcher has put it:

Even the most advanced theoretical models, generated by research, are not yet adequate to account, in a precise manner, for the mental processes of a pupil in a classroom situation and the exact use he or she makes of feedback (Perrenoud 1998 p.87⁵)

Yet some experienced teachers do this well, albeit intuitively. As Sadler (1998) put it:

Teachers bring evaluative skill or expertise in having made judgments about student efforts on similar tasks in the past. [...] In non-convergent learning environments, this automatically exposes teachers to a wide variety of ways in which students approach problem solving, and how they argue, evaluate, create, analyse and synthesise. (p. 81)⁶

The task of setting up a systematic guide to underpin this skill or expertise would be a formidable enterprise.

However, a commitment to down-grade rote learning and prioritise teaching for understanding gives rise to other challenges. As Perrenoud puts it 'it is important to distinguish the regulation of on-going activities from the regulation of the learning process' (p.87). The choice, even within a commitment to improving learning, is either to steer the classroom dialogue to reach the target of understanding concept X, or to work towards a longer-term target of developing one or more of a set of reasoning skills, concept X being treated simply as a vehicle for the exercise which can be set aside if the reasoning work requires this (an extreme example is the Cognitive Acceleration in Science Education scheme, in which teachers have to devote one of their science lessons each fortnight over two years to the scheme's activities: in these lessons they will not be able to 'cover' the syllabus). Such choices are for the teacher to make - there can be no simple rule.⁷

2.6 *Strategic overview*: The planning by a teacher has to pay attention to several choices. One is to strike a balance, and optimise the interactions, between the opportunities for

⁵ Perrenoud, P. (1998) From formative evaluation to a controlled regulation of learning processes. Towards a wider conceptual field. *Assessment in Education: Principles, Policy and Practice*, **5**(1), 85-102.

⁶ Sadler, D.R. (1998) Formative Assessment: revisiting the territory. *Assessment in Education-Education*, **5**(1), 77-84.

⁷ Shayer, M. & Adey, P. (eds.) (2002) *-Learning intelligence: cognitive acceleration across the curriculum 5 to 15 years*. Milton Keynes: Open University Press. This issue is explored in some detail in Black and Wiliam (2009)

dialogue afforded by classroom work and those afforded by written work. The latter also involves the choice of task and then the formative choice of the most appropriate form of feedback, given that such feedback can be made without the immediate urgency of the classroom, and that it can be tuned differently according to the different needs of each individual learner.

Another of the choices to be made is between whole-class discussions, where special techniques have to be deployed to secure the active participation of as many pupils as possible, and discussion within and between sub-groups. For the latter, there is well-documented evidence that small-group work can be ineffective unless special training is given to ensure that all participate, that each group keeps on-track, and that the discourse is rational and collaborative rather than assertive and combative.⁸ However, this literature does not give clear guidance about optimum ways of switching back and forth between whole-class and small-group work.

A third strategic choice is to choose a stage at which a review of progress would be profitable. Such review interrupts the flow and takes up time, but it can help both to find evidence of the developed understanding of every individual pupil in a way that is hard to collect in either whole class or group discussions, and also to give feedback to the teacher about the overall successes and failures of the work, so guiding the choice of the next steps in the learning programme. Viewed only in these terms, some activities which may be labelled as summative assessment could have a solely formative purpose. It is only when other purposes, which may use the same review evidence, are to be served, that the complexity of the interplay between formative and summative becomes problematic.

Recommendation One: The proposals in this section should be used to further develop the Council's Reflection Framework in order to refine the model of pedagogy, to clarify the role of formative assessment within that framework, and to clarify the relationship between the Principles and the Practice of formative assessment within that model.

3 Summative Assessment: coherence and clarity about purposes

3.1 Three levels of summative: In considering the purpose for which reviews of evidence are undertaken, it is helpful to think in terms of three levels of purpose, the formative level, the school-level and the accountability-level. At the formative level, as considered above, reviews are needed by both teachers and learners as occasions to reflect on progress. Summative tests (or more properly, tests designed primarily to serve a summative function) provide ways of eliciting evidence of student achievement, and used appropriately, can prompt feedback that moves learning forward. They can also communicate to learners what is and is not valued in a particular discipline, thus communicating criteria for success. Where this has been done, it opens up the possibility of students helping one another, and using the tests as a guide to planning their own reviews, both during preparation, and afterwards in

⁸ See (a) Mercer, N., Dawes, L., Wegerif, R. and Sams, C. (2004) Reasoning as a scientist: ways of helping children to use language to learn science. *British Educational Research Journal*, **30**(3), 359-377, and (b) Blatchford, P., Baines, E., Rubie-Davies, C., Bassett, P., & Chowne, A. (2006) The effect of a new approach to group-work on pupil-pupil and teacher-pupil interaction. *Journal of Educational Psychology*, **98**, 750-765.

involvement in the marking process⁹.

At the school level, less frequent reviews are needed when teachers or parents or the school management have to take decisions about a pupil's future e.g. for future choice of subjects, for forwarding information to next year's teacher, and so on. Internally generated summative assessments are essential to this purpose, and since pupils' futures may be affected by the results, it is important that they be carried out with a high level of skill. In this level of summative, schools have full responsibility and control. The limitations, on the extent and the types of evidence that can be collected, within constraints of time and of cost, are less severe than those for externally administered testing: so, for example, use can be made of pupils' reports on an extended project as part of a personal assessment portfolio. The problems here are (a) teachers need to develop skills in formulating and auditing assessment exercises and (b) some evidence of comparability, across teachers in the same department, and between departments in the school, is needed. It seems likely that many teachers and schools will need substantial professional development in developing and handling both the instruments and procedures to serve this purpose. At this level therefore, the individual teacher, the school subject department, and the school management share responsibility, and the assessment occasion is more significant, and can be more stressful, than similar exercises at the first level.

At the third level the assessments have to provide for local to national policy requirements for accountability. Insofar as schools' assessments at the school level may be used for this purpose, some degree of alignment between the tasks used both within and across schools is needed if the moderation (for comparability) of samples of pupils' work is to be feasible. It is here also that the issue of standards has impact. Such moderation, as well as serving the wider accountability purpose, can also have value in improving the school-level summative work: the professional development that is recognised in the Highland's LSGs is evidence of this second type of benefit.

3.2 *Tensions between the levels:* Two types of tension become evident here. One is between the different requirements of the different levels. The second is the tension between the formative and the summative purposes, which arises at all three levels. I suggest that confusions arise if these two are considered together rather than separately.

For the different requirements of the different levels, there is first the choice, at each level, between teachers' summative assessments and those conducted or directed from outside the school. The issue of the dependability of these two approaches is central here, and this involves both validity, and one of the components thereof, reliability. Formal external tests are trusted as reliable: the evidence however is that this trust is misplaced. There are many threats to reliability: some, notably inconsistency of marking, can be reduced by careful procedures. Other, notably variations of pupils' performances between one testing date and another, and between one version of the test and another, cannot be avoided. There is good evidence that in public examinations in England, between 10% and 30% of candidates are not

⁹ A 2009 study, with a small sample of Highlands teachers, for the SQA, presents many impressive examples of ways in which teacher have done this (Hayward,L.,Boyd, B.,McBride, G. & Spencer, E.(2009) SQA in preparation for publication).

given the level they deserve¹⁰. The limitation here is inescapable, for a main source is that in a timed formal examination, only a small sample of all of a candidate's possible achievements can be assessed; one estimate shows that to reduce the misclassification to about 1% of candidates would require about 30 hours of testing. Teachers' assessments can range over longer periods and over many occasions, so that they could in principle do far better.

Validity is a broader and more complex issue. Put simply, the issue is whether a pupil, or a parent, or an employer, can assume that a high grade in (say) maths. means that the pupil is good at maths. This immediately raises the question of what it means to be good at maths, and there may not be consensus on this issue. If it means ability to respond to short, closed, highly structured problems calling for operation of well-defined algorithms, formal tests can produce relevant evidence. If it means to tackle an everyday problem by selecting and deploying appropriate maths methods, then any alternative to in-course assessment by teachers seems impossible: because of the time required to tackle any one problem, the need to explore several such problems to achieve reliability, and the fact that many such problems are tackled in adult life by collaborative discussion,

There is a more subtle issue here: there is good evidence that if a group of pupils with the same initial test result are all, after the test, given the same help and then tested again, they will then perform quite differently from one another. If 'being good at X' involves the ability to take advantage of help in order to improve, then the performance with help may be a more valid indicator than the performance without it. Public examinations have not even tried to examine this possibility, known in the literature as dynamic summative assessment; teachers' could use this approach¹¹. One of reasons why, in other contexts, interviews are seen as essential in the selection of candidates, is that in interviews a dynamic approach can be adopted.

3.3 Problems with teachers' summative assessments: Whilst the potential values of teachers' summative assessments in contributing at the accountability level can be argued as in the above section, several outstanding weaknesses undermine confidence amongst some, whether in the general public or in politics, that these values can be achieved. Such sceptics argue that teacher assessments lack comparability between different teachers and between schools, may not be aligned with national standards, and may be flawed because of the difficulty of checking, that work is the candidate's own and that teachers have not been tempted to inflate any of the scores.

¹⁰ Black, P. & Wiliam, D. The reliability of assessments pp 214-239 in John Gardner (ed.) *Assessment and Learning*. London: Sage, See also pps. 22-26 in *House of Commons , Children, Schools and Families Committee: Testing and Assessment. Third Report of Session 2007-2008 Vol.I*, and Black, P., Gardner, J. and Wiliam, D. (2008) Joint memorandum on Reliability of Assessments submitted to the Committee in *Volume II. HC169-II*. Norwich: The Stationery Office. Ev. 202-5. ISBN 978 0 215 52041 8. Research into this issue has been taken on board as one of its priorities by OFQUAL, the new regulatory body in England,

¹¹ See chapter 4 in Poehner, M.E. (2008) *Dynamic Assessment*. Springer, and also Ahmed, A. & Pollitt, A (2004) *Quantifying support: Grading achievement with the Support Model* University of Cambridge Local Examinations Syndicate. Paper presented at the IAEA Conference, Philadelphia, June 2004.

It is important to recognise these objections, and to set up procedures convince the sceptics that their fears are unjustified. The problem is to do this in ways that are helpful to good learning and thereby acceptable to teachers. For example, in one aspect of the SQA approach, already moderated assessment exercises are available on a central 'bank'. A recent evaluation study commissioned by the SQA shows that a few teachers perceive the use of this bank to be an SQA requirement - which it is not - and find this perceived requirement oppressive and regard it as counter-productive. The development of assessments by ASGs could encourage greater teacher ownership of the system, align with SQA policy, and point to an alternative way to assuring comparability, by implementing robust inter-school moderation procedures.

Any policy to strengthen this aspect of teachers' work should attend to several obstacles¹². The long-standing dominance of external requirements, backed by the perceived threats of accountability pressures, may have de-skilled many teachers. For them, it has become far easier to rely on externally provided instruments, than to invest the effort and face the risks involved in taking responsibility to explore the potential of providing more valid evidence from activities in their own classrooms. Such brave ventures have to navigate the pressures of requirements of both national and local government, of inspectorate findings, of the trust of parents, and of departmental and whole-school policies. Such navigation is very demanding if these requirements are not mutually consistent.

To develop their potential, teachers need to have clear understanding of the concepts of validity and reliability as they apply in their own subject areas: hitherto, these topics have been given little priority in initial and in-service training. Teachers can develop their concept of validity for this calls for reflection on their own reasons for believing in the educational values of their subjects. However, to many an exploration of the reliability component which goes beyond assuring consistency in marking to confront the issues of error in grading which arise from limited sampling, or which looks into different interpretations of standards, seems to be a technical exercise of little relevance to them, so that they are less ready to participate. Yet if the conditions for effective CPD so clearly spelt out in the Highland Council's reports, i.e. stimulus from external sources and strong mutual support between colleagues as they develop new innovations in their practice, are met, then improvements can be secured.

A promising approach is to develop portfolio collections of a set of samples of pupils' work, with each pupil sharing responsibility for the production of his/her own. Each portfolio can include results of formal tests, assignments taken under controlled conditions, and work produced under the less formally controlled conditions that open-ended projects require. Crucial to success here is the quality of the questions or tasks that are used, and in this respect many teachers will have to develop new insights and skills, and to learn as they proceed (e.g. when it turns out that a carefully formulated task does not challenge 20% of the class, or completely defeats a different 20%, or when the responses fail to reveal evidence of the understanding which the task was designed to assess). The 2009 SQA evaluation shows how inventive teacher groups can be in making good use of summative opportunities when

¹² Here I draw on a recent study at King's College, in collaboration with 3 schools: see Black, Harrison, Hodgen, Marshall: Validity in Teachers' Summative Assessments, *Assessment in Education* (under review).

working within a set of external requirements: such talent would surely be productive if in any future system, groups were given the responsibility, within any broad framework of national requirements, of producing robust evidence of comparability within alignment to standards.

3.4 Moderation: Moderation procedures are an essential part of any such enterprise. These start at the school level as within-subject and across school procedures, so each school's own policy should be clearly specified and provide an effective framework to secure the comparability of high-quality summative judgments. These would involve, for example, blind marking¹³ of samples of work so that differences in judgment are revealed and can then be explored, and some compromise, between imposing use of the same tasks by all teachers of the same subject at the same level, and allowing for some to be at the choice of each teacher to allow for different personal approaches. These stringencies then set a basis for a similarly robust approach, again involving blind marking and some common tasks, for inter-school moderation exercises; here the rules of procedure should be established under guidance from the local authority. This two-level approach is expensive in teachers' time, but those who have been involved in such procedures have found them to be fruitful for their professional development and not merely checking-up exercises.

3.5 The formative-summative interface for within school summative assessment: Tension between formative (FA) and summative assessment (SA) is commonly identified as a problem, but I suggest that it cannot be resolved without breaking it down into two parts. The **first part** is the tension between FA and SA in those exercises in SA for which teachers have total control. Where this problem cannot be resolved within schools, it is likely that the external SA will dominate the internal SA and the tension will be seen as between the school and the political will to control. Yet it is for the profession, within those domains under its own control, to first develop the solution to the problem. There are many possibilities here: e.g. it could be argued that in a classroom based coursework task (e.g. an open-ended problem in maths, or composing a poster to advertise an event in English) the SA will be 'unfair' if pupils are given help. But, as explained above, there is a counter-argument, for the ability to take advantage of help in order to improve ought to be assessed. There is no 'solution' offered here – merely some ideas as to the directions in which school-based development projects might be the way in which to develop new approaches to the problem. Sections 3.2, 3.3 and 3.4 above indicate some of the ideas which such projects might explore. However, schools and inter-school groups cannot be expected to invest effort in such exploration unless they are set free from at least some of the constraints of the present system.

¹³ Blind-marking is a procedure in which samples of work are copied and distributed to a group without indication on them of any marking that may have taken place. Each member then marks the work independently, and records in writing their marks. These written records are collected and exchanged at the start of the moderation meeting, and then reviewed to highlight problematic cases which call for discussion. The initial records also give some indication of the degree of comparability already achieved. A few teachers object to such procedures because they think they are not being trusted - whereas the point is to help them to give the public evidence that their decisions are trustworthy

Recommendation Two: A development project should be conducted with and within schools with the aim of auditing and then improving the quality of their internal summative assessments

3.6 *The formative-summative interface and assessment for accountability:* The ***second part*** is the tensions created between the requirements on national or local assessments and teachers' own practices. The accountability pressures that are exerted by such requirements will influence classroom practice more powerfully than any resolution of the FA/SA tensions developed by schools themselves. Since these external procedures have not been developed in the light of evidence about ways to resolve the tensions, teachers are left with having to cope with inconsistency across the FA/SA interface. The effects of this on classroom practices have been well documented.

Such difficulty may well be inevitable where the formulation of national requirements and procedures has been 'top-down' i.e. designed from outside rather than evolved by growth from inside, so that they lack ecological validity. The involvement of teachers in such designs is no guarantee, for in a consultative context they may wish to object, but they will usually lack any basis in experience to support arguments for alternatives. Those responsible for a national initiative should set the broad conditions which any policy change should meet, but should leave it to the professionals to work out ways to meet these, recognising that such developments take time, require support, and may involve development and evaluation of alternative approaches before choice are made.

3.6 *Accountability:* The evidence of the recent past is that if accountability is meant to improve the quality of teaching and learning, then it has failed. Any attempt to improve by applying high-stakes indicators is bound by the limitation of Goodhart's Law, which states that "All performance indicators lose their meaning when adopted as policy targets". Examples of this abound, e.g. Railtrack's performance targets, and National Health Service waiting list targets. National school achievement targets are no exception: in England, they did lead to gains in the first few years, but these have then levelled out - a well-known effect in the USA as teachers learn to teach to any new test - whilst international tests do not reflect this trend, Shayer's replication of his science reasoning tasks show that national performance on these has declined over the same years¹⁴, and two international surveys have given our country a low ranking in assessments of the happiness and confidence of school children.

There is no easy solution to this problem. Setting clear criteria for the methods and procedures by which pupils are advised or judged is a national responsibility: so schools should be accountable for the way in which they fulfil their assessment responsibilities. To say they are responsible for the results goes beyond this. The extent to which factors within a school's control account for the variability in assessment attainments between schools is very low - one estimate puts it below 10%. The developments in England first of value added indices, and then of context-value-added (CVA), have struggled with this problem. As such refinements are criticised and sharpened, they become so complex to understand that the public are confused or mis-led, and they still leave the system open to manipulation (accept weak students to lower intake scores, and then find ways to get them to leave or transfer en

¹⁴ Shayer, M. (2006) Thirty years on - a large anti-Flynn effect? The Piagetian test volume and heaviness norms 1975-2003. *British Journal of Educational Psychology*, **77**, 25-41.

route to raise the outcome scores). A light touch system, in which each school produces a comprehensive account of all the ways in which it cares for children, and in which extraction by the media of data which could be presented in simplistic league tables is made more difficult, would seem a way forward. No set of changes would be proof against Goodhart's Law, but they could make the system far closer to the ideal, which is that attempts to enhance performance on the indicators actually improves the quality of children's education.

Recommendation Three: The national policy for ensuring the comparability and validity of the assessments made by teachers should be examined in the light of the recent evaluation of its effects in the Highland Council's schools, with a view to supporting improvements in the quality of teachers' summative assessments and to reformulating the accountability system so that it gives better support to sound learning practices.

4 Support for teachers in change

4.1 Professional development for change: The principles involved here were clearly stated in 1998 by Black & Wiliam in *Inside the Black Box*¹⁵:

Thus the improvement of formative assessment cannot be a simple matter. There is no 'quick fix' that can be added to existing practice with promise of rapid reward. On the contrary, if the substantial rewards of which the evidence holds out promise are to be secured, this will only come about if each teacher finds his or her own ways of incorporating the lessons and ideas that are set out above into her or his own patterns of classroom work. This can only happen relatively slowly, and through sustained programmes of professional development and support. This does not weaken the message here—indeed, it should be a sign of its authenticity, for lasting and fundamental improvements in teaching and learning can only happen in this way. A recent international study of innovation and change in education encompassing twenty-three projects in thirteen member countries of the OECD has arrived at exactly the same message in framing advice to the member countries about effective policies for change (Black and Atkin 1996¹⁶). (p.15)

There is little to be added here, for the Highland Council reports show their strong commitment to the involvement of teachers in fashioning change. What might be added is that teachers who are supported in carrying out such changes should be willing to justify such support, and to help other teachers who might have to follow the same path, by attending to the need to produce evidence, of a variety of kinds and at more than the simply anecdotal level, of the benefits they have secured.

Recommendation Four: The central role that teachers should play in developing or reformulating proposals for change should be recognised by sharing with teachers the task of developing new policy proposals, so that such proposals are both feasible within good classroom practices, and positively support such practices.

¹⁵ Black, P. & Wiliam, D. (1998). *Inside the black box: raising standards through classroom assessment*. London: GL Assessment.

¹⁶ Black, P. & Atkin, M.(eds.) (1996) *Changing the Subject*. London Routledge – for the OECD.

4.2 *The challenges of change:* The adoption of ideas about formative assessment has revealed many examples of groups who see it, from their superficial examination, as requiring teachers merely to include a few new tactics to their existing style of teaching. Anyone who has worked seriously to achieve the benefits of its adoption will know that this is a travesty. What is called for is rather, for many teachers, is a change of role, a change in their perception of what it means to be an effective teacher. The following quotations illustrate this¹⁷:

Actually thinking about teaching has meant that I have been able to come up with ideas and strategies to cope with whatever has arisen and has contributed greatly to my professional development’
Susan, Waterford School p.19

There was a definite transition at some point, from focusing on what I was putting into the process, to what the students were contributing. It became obvious that one way to make a significant sustainable change was to get the students doing more of the thinking.

Tom, Riverside School p.20

What formative assessment has done for me is made me focus less on myself but more on the children. I have had the confidence to empower the students to take it forward

Robert, Two Bishops’ School p.22

Such changes were only achieved near the end of a two year project to support teachers in making formative ideas work in practice: during that time many of the teachers described the changes as ‘pretty scary’, for they felt they were giving away their previous means of keeping control with pupils. The reasons for such difficulties can be seen from the accounts given in section 2 above, where the various challenges posed by adopting a new approach to pedagogy are spelt out. The difficult and slow pace of deep change is a familiar finding of those who have worked with teachers on development of other innovations in practice. This feature should be borne in mind when any such changes are proposed.

4.3 *Support from research:* Whilst there should be scope for some research in education which ignores any potential for practical application, more of research effort should be aimed at developing the potential for improvement through innovation. However, some research designed to inform policy may not validly do so because, despite rigour in the measurement and analysis of relevant variables, its focus on readily measurable factors, such as pupils’ scores, or class size, or the test outcomes from different curricula, may lead to proposals which may not be based on an understanding of the links in practice between cause and effect.

The research that can be a basis for policy change must be work which explores how teachers have used ideas generated by research, adapted them, and by making them their own established new knowledge which works in their practice and which can therefore be basis for a policy aiming at widespread adoption

Recommendation Five: The research community should be encouraged and supported both to propose, on the basis of research findings, innovations which deserve to be the subject of trials, and to work in close collaboration with teachers and schools in developing

¹⁷ Black, P., Harrison, C., Lee, C., Marshall, B. & Wiliam, D. (2002). *Working inside the black box: assessment for learning in the classroom*. London: GL Assessment

new working practices through the conduct of such trials and in carrying out evaluation of their outcomes.

5 Coherent systemic reform: alignment of policies and agencies

The existence of national agencies which are separately responsible for different components of the system makes coherence hard to achieve. A notable example is that curriculum authors often set out and promulgate grand and worthy aims, but leave it to a separate examination agency to operationalise these in actual instruments and procedures. Neither agency may pay close attention to the implications for classroom practice.

Teachers are then left to work out how to teach to the grand aims whilst at the same time preparing pupils to cope with the examiner's instruments, having in mind that the public and the policy makers will use the outcomes to make schools accountable. Teachers have to resolve any neglected requirements, inconsistencies and contradictions which might emerge from changes, which have not been developed with conscious and careful attention to systemic coherence, it should be routinely expected that no curriculum aim can be approved unless its attainment can be exemplified in some sample, or collection of samples, of performance by pupils, and unless evidence produced that changing teaching to achieve such performance is both practicable and productive in the improved learning of pupils.

Any policy for change should also take into account the need to match new policies to the capacity of the teaching community to implement the changes involved. **If** such policies have been developed in, and tested against, the complex realities of classroom and school life, implementation is likely to be hard slow work; **if** they have not been so developed they are likely to fail.

Recommendation Six: Agencies which contribute to policy formulation should, before any promulgation of their policies, work together to ensure their coherence between their likely effects on curriculum, assessment and pedagogy, and to produce evidence that policies will, when applied within the working conditions of teachers, produce their intended effects.

6 Ways forward

The six recommendations set out above may be briefly summarised as follows:

- 1 Further develop the model of pedagogy to clarify and strengthen its several components, notably those involving assessment.
- 2 Improve schools' own summative assessment policies and practices.
- 3 Match such improvements to improved ways of using schools' own assessment of pupils within accountability requirements that give more support to classroom learning.
- 4 Recognise the central role that teachers should play in developing and refining new policy initiatives.
- 5 Encourage researchers to participate, in partnerships of collaboration with teachers and schools, in the formulation and trial of potentially valuable innovations.

- 6 Seek ways to enhance collaboration between national agencies in the development of new policies so that teachers are presented with plans which are mutually supportive in their effects on curriculum, assessment and pedagogy.

What may stand out in any overview of these is that recommendation 4 must be an important aspect of attempts to meet any of the other five.