

Policy and New Products

# Research Report 11



## **Literature review on internal assessment for the National Qualifications for the Future project**

## Background

This report was written by Rob Van Krieken, on behalf of SQA in September 2006.

A large part of the literature on the use of internal or teacher assessment versus external assessment is produced by the Assessment Reform Group. This group is mostly oriented on England — the situation in Scotland is different in the following respects:

- ◆ England has obligatory Key-stage tests at ages 7, 11, and 14, while Scotland has an optional use of 5–14 assessments (currently English and Maths, other test material will become available in the future).
- ◆ The results of Key-stage tests in English schools are published in league tables, and competition between schools is very strong. This exacerbates the usual tendency to teach to the test. This pressure does not exist in Scotland until the exams in S4. Instead, in Scotland there is the national Scottish Survey of Achievement, which does not report on individual pupils or schools (but is starting to report by authority).

American literature needs to be read against a background of intensive use and preparation for multiple choice, and mostly knowledge-based, external tests, which in no way can be considered to be good formative practice. This situation is caused by a strong focus on accountability by results, and a lack of generally accepted federal or national exams (Wolf, 1996).

This explains why the Assessment Reform Group has drawn attention to the negative effects of external exams and has been looking in its literature reviews for possible advantages of internal assessment. Many arguments are indirectly directed at key-stage testing. There is little acknowledgement of the reasons why tests and exams are needed and used.

There is no doubt however, that:

- ◆ Teachers are not trained well in assessment. This situation is reported in many countries, including England and Scotland. Reports of the improvements made by introducing communication with pupils implicitly acknowledge the poor state of assessment beforehand (Lingard, Mills & Hayes, 2006; and Hayward, Spencer & Simpson, 2005).

- ◆ Teachers tend to use exams as the most practical embodiment of objectives or standards. To what extent they teach to them depends on the status of the exam, and the use of exam results, as well as the availability of other teaching material and examples of good practice.

## Internal assessments

Internal assessments can serve two different purposes. The first ('formative assessment') is to help students to find out what they still should do and how. The second ('summative assessment') is to gather information to inform others about the progress a student has made so far. It is generally acknowledged that teachers are not well trained in formative assessment (Hayward, Spencer & Simpson, 2005).

The high status and ubiquity of summative assessment in school administration, as well as in external exams, has led teachers to think that assessing and reporting in the way exams or external tests do is the only way learners can and should be assessed (Assessment and Learning Research Synthesis Group, *A systematic review of the evidence of the impact on students, teachers and the curriculum of the process of using assessment by teachers for summative purposes*. EPPI-centre, 2004).

## Internal assessment as part of external exams

In some cases, summative internal assessment is part of the examination. Evidence about the quality of teachers' own summative assessments is mixed (Wilmot, 2005), but improving (Dhillon, 2005). Even though teacher-assessed course components do not necessarily assess the same as external components, teachers mostly rank their candidates in the same order as external assessments do. However, their marks tend to be consistently higher or consistently lower.

Assessments by teachers are said to be better than external assessments for the following reasons:

- ◆ They provide better learning. (This assumes that summative evidence can be derived from evidence collected for formative purposes, so formative assessment is the primary focus. The argument is that good teaching requires formative assessment. Internal assessment as part of an exam does not produce good teaching or formative assessment, but it at least it doesn't discourage it as much as external assessment.)
- ◆ They can assess the whole curriculum, including higher-order skills. (This assumes that external assessments cannot or do not assess higher order skills such as reasoning, application of knowledge to new contexts, problem solving, or evaluating; and that assessments by teachers will, and will be of acceptable quality.)
- ◆ They can assess the whole curriculum, including practical skills. (This assumes that there is no model for externally assessing practical skills such as cooking, music performance, or speaking or listening; and that that teachers' practical assessments will be of acceptable quality.)

- ◆ They are needed because central external exams can assess only a small sample of authentic problems, and would be insufficiently reliable on their own.

Each of these reasons is discussed in the following paragraphs.

## **Better learning**

The ‘Assessment is for Learning (AifL)’ initiative has provided guidelines and support for formative assessment as a means to gather information and provide meaningful feedback to the learners. The evidence shows that it leads to teaching with better results (James, Black, & McCormick, 2003).

‘Better results’ are defined here as ‘deeper learning’ (ie better integration of learning within existing knowledge and skills, as opposed to ‘rote learning’), and learning of higher-order skills or general and transferable skills (Gipps, 1996).

It is also possible to prepare better and more efficiently for summative assessments by using many of the techniques that are successful in formative teaching. This is not because these techniques in themselves are specifically formative or summative, or because they have anything to do with assessment — they are simply good teaching practice. Essential elements are:

- ◆ Making sure that learners know ‘the standard’, ie what they are supposed to learn (eg by analysing exam questions).
- ◆ Making sure that learners can identify the difference between the standard and their own performance.
- ◆ Giving meaningful feedback (eg indicating how learners can improve, instead of just reporting a grade or mark).
- ◆ Using peer-teaching (peers tend to understand each other’s problems better and use language that is less abstract and better understood by their classmates).

From experiences so far, it is abundantly clear that improving teaching by filling the existing gaps in teachers’ training and practice in assessment is an extensive, expensive, and extended, but essential, process.

## **Higher-order skills**

In Scotland, the Standard Grade exam consists of usually three elements, one of which assesses knowledge and understanding, the second application and analysis or evaluation, and the third practical skills. The second element, (the higher-order skills of application, analysis and evaluation) is examined and marked externally. National Courses have their (often increasing) share of higher-order questions. This proves that it is possible to assess higher-order skills externally.

The question remains whether teachers could write formative assessments of higher-order skills and use these to teach learners better. Experiences in England and Queensland were disappointing (Bullock et al, 2002, Lingard, Mills and Hayes, 2006). The current practice of setting a prelim and marking or grading this according to exam

guidelines months before the exam, shaped by SQA's requests for estimates, has its weaknesses in its use for summative purposes as well as for teaching. (Howieson & Hart, 2004).

## **Practical skills**

Both in England and in Scotland, exams have summative 'course components' or 'internal components' which mean that students produce evidence during the course, which is assessed by their teacher (and moderated). In England, these suffer from low quality assessment, or from very detailed guidelines for the selection and assessment of tasks (Bullock, Bisshop, and Reid, 2002), or they are suspect because of high results and assumed lack of authenticity (caused by teacher instructions, revising, parental support, and the availability of 'off-the-shelf' products). Also, teachers interpret guidelines differently (Yung, 2002).

In Scotland, at secondary level, there is a systematic use of internal assessment in Standard Grade, where one of the usual three elements tends to be more practical and is assessed internally. National Courses all have internally-assessed units (usually three) which assess whether a minimum level has been achieved. The more practical courses, such as music, hospitality, and languages, have practical components which are performed in school; and other courses have dissertations, investigations, and folios etc, which are produced in school, under supervision and with guidance from the teacher. Most of these, however, are marked by examiners, whether visiting or centrally. Project Based National Courses (PBNCs) are mostly externally moderated in their development and/or evaluation stage.

It is indeed essential to the use of these components that the teacher is capable of switching from a supporting role during the production of the evidence (drafts, design etc), to the role of examiner when selecting final evidence (eg the best attempt, or a product produced independently and not revised according to instructions — see Griffith, 2004).

## **Better assessment**

The argument that teachers could assess the whole curriculum better than external assessments rests on a number of assumptions. The first is that teachers' assessments can be of sufficient quality.

Gibbs (1996) suggests that if teacher assessment cannot be reliable, at the very least it can be comparable. (The more statistically minded will counter that comparison of unreliable results cannot lead to any reliable conclusions.) Many authors assume that teachers will be able to produce summative assessments or judgements from evidence gathered for formative purposes, but there is little evidence of that (Black, 2004).

The second main assumption is that teachers will have time. This seems to rest on the assumption that external assessments or exams will disappear altogether.

Wolf (1996) warns that it is too easy to attribute higher motivation in other countries only to internal assessment, when there are so many other factors:

- ◆ Higher motivation in Germany and France among weaker pupils can be explained by the loss of important chances if they perform badly in their teacher's assessments compared to the relative unimportance of these for these pupils in England and the USA.
- ◆ The effect of assessment types must be understood within the whole structure of the labour market.
- ◆ There are few 'general' relationships between type of assessment and student motivation, or between assessment methods and control systems.

## **External assessments**

There are good reasons to have external assessments, such as the need for certification, comparability, and system feedback. This is why many countries are introducing them (West & Crighton, 1999). The question is whether there are any alternatives or ways to prevent or diminish unwanted side-effects. Black (2004) says that summative tests need not be harmful and can be a positive part of the learning process.

Exams or external assessments are instruments, and so have a standard error of measurement, which indicates how accurate the reported score is. A poor ratio of assessment time to content and skills covered can diminish the reliability of assessment, particularly when content or skills are not homogeneous. This is the case when authentic contexts and situations are used.

There are countries without final exams where certification is or was based on teacher judgements (Finland, Sweden and eastern European countries). University entrance, however, still required an entrance exam. Sweden has mandatory national tests at the end of compulsory education to moderate teachers' judgements against national standards. Eastern European countries (Poland and Slovenia) are increasingly introducing national assessments to provide comparable information about educational results as feedback on their education system (West & Crighton, 1999; Kellaghan, 2004). In Greece, the low status of state certificates (and the small number of university places available) has led to a host of private language schools preparing students for American or Canadian tests providing entrance to university.

## **Out of the fire ... and clear of the frying pan**

Clearly, internal assessment and external exams should not be seen as alternatives, but as instruments that need to be used with care, and for the purpose they serve best. This means that the various functions performed by each need to be analysed and disentangled, for instance as:

<b>Function</b>	<b>Assessment method and locus</b>
feedback for formative purposes	→ internal assessment <sup>1</sup>
summative progress reports within school	→ internal assessment <sup>2</sup>
information supporting progression advice	→ internal assessment and optional tests <sup>3</sup>
certification	→ both internal and external assessment <sup>4</sup>
university admission	→ based on certification <sup>5</sup>
system feedback	→ (inter)national surveys <sup>6</sup>
school accountability	→ HMI <sup>7</sup>

<sup>1</sup> It is clear that improvements in their quality are needed and possible (Hayward, Spencer & Simpson, 2005).

<sup>2</sup> Best practice examples and development are needed to minimise the frequency of summative testing and maximise the quality. Provision of optional assessment material as for 5–14, and through NABs (instead of an exam) might contribute.

<sup>3</sup> This means certification only when leaving school. In view of the early stage, certification of a core curriculum and cross subject skills and experiences might call for a group certificate rather than individual subject certificates. NABs might be made available for optional use to support progression decisions.

<sup>4</sup> There are obvious advantages in teachers conducting summative assessment. Wiliam (2000) suggests more, but standardised and moderated, teacher assessment to achieve a valid and reliable total assessment. Lengthy teaching to the exam might be avoided by weighting internal components heavier than the external component, while still using the external component to moderate.

<sup>5</sup> As currently.

<sup>6</sup> Currently, Scotland takes several surveys (SSA, PISA, TIMMS, PIRLS) though most of these are not on a yearly basis. Clarity about the level of feedback is necessary (national, EA, school or department). A survey could provide a more extensive picture than a certification assessment because it can spread all contents and skills systematically over all pupils. Also, as long as it reports at authority or national level, there would be no incentive to teach to it.

<sup>7</sup> HMI reports have high status and are comprehensive. They provide better founded information on specific schools than league tables produced by newspapers, which concentrate on academic results.

## References

- Assessment and Learning Research Synthesis Group, *A systematic review of the evidence of the impact on students, teachers and the curriculum of the process of using assessment by teachers for summative purposes*, EPPI-centre 2004
- Assessment & Learning Research Group, *A systematic review of the evidence of reliability and validity of assessment by teachers used for summative purposes*, EPPI Centre 2004
- Black, P (2004), Issues in Assessment by Teachers, Appendix E in Report of ASF Seminar 1, Cambridge, 2004
- Broadfoot, P & Black, P (2004) Redefining assessment? The first ten years of Assessment in Education, *Assessment in Education*, 11 (1)
- Bullock, K et al. (2002) Learning from coursework in English and Geography, *Cambridge Journal of Education*, 32 (3)
- Harlen, W & Deakin Crick, R (2003) Testing and Motivation for Learning, *Assessment in Education*, 10(2)
- Gipps, C (1996), Assessment for learning, in: Little, A & Wolf, A, *Assessment in transition. Learning, monitoring and selection in international perspective*
- Griffith, S A (2004), *Assuring Fairness In School Based Assessment: Mapping The Boundaries Of Teacher Involvement*, Paper presented at the 31<sup>st</sup> Annual Conference of the International Association for Educational Assessment (IAEA)
- James, M, Black, P, & McCormick, R (2003) *Deepening capacity through innovative research design: researching learning how to learn in classrooms, schools and networks*, Paper AERA
- Hayward, L, Spencer, E & Simpson, M (2005) *Assessment Is For Learning: Exploring Programme Success*
- Howieson, C & Hart, J (2004) *How Do Centres Estimate Pupils' Attainment?* Report to the Scottish Qualifications Authority
- Lingard, B, Mills, M and Hayes, D (2006) Enabling and aligning assessment for learning: some research and policy lessons from Queensland, *International Studies in Sociology of Education*, 16 (2) 83–103
- Somerset, A (1996) Examinations and educational quality, in Little, A & Wolf, A, *Assessment in transition*, Pergamon
- West, R & Crighton, J (1999) Examination Reform in Central and Eastern Europe: issues and trends, *Assessment in Education*, 6 (2)

Wiliam, D (2001), An overview of the relationship between assessment and the curriculum, in D. Scott, *Curriculum and Assessment*, Ablex publishing

Wiliam, D (2001) Reliability, validity, and all that jazz, *Education*, 23 (3), 17-21

Wilmot, J (2005) Experiences of summative teacher assessment in the UK. A review conducted for QCA

Wolf, A (1996) Individual choices, incentives and control: understanding assessment dilemmas, in Little, A & Wolf, A, *Assessment in transition*, Pergamon

Yung, B (2002) Same assessment, different practice, *Assessment in Education*, 9, 97-117