

Policy and New Products

# Research Report 14



**Internal Assessment, External Assessment, and  
Assessment for Learning — a think piece for those  
involved in developing assessments**

# Introduction

## Purpose

This paper was commissioned in May 2007 to explore the ways in which externally developed assessments can contribute to teaching for learning in the classroom and how external assessment should change to be best aligned with Assessment for Learning.

It may be helpful to start with a brief description of different types of tests and their use, as they developed in the history of education. For full descriptions we refer to existing handbooks. This paper selects only what is relevant to issues discussed recently in the context of Curriculum for Excellence and the new generation of Qualifications.

## Background

SQA fully accepts the drive to improve teaching by ensuring that assessment should serve the purpose of teaching, and that teaching should not be affected negatively by assessments that are put in place for accountability or certification.

Successful teaching is based on evidence, produced by learners during learning activities and in assessments, that demonstrates whether:

- ◆ a learner has learned, understood and can apply in different situations
- ◆ a learner is ready to learn
- ◆ despite attempts to learn and teach, a learner has not yet understood and demonstrated achievement
- ◆ learning and teaching have not yet been successful (and what can be done to improve this)

There is general agreement that evidence is best produced in an interactive approach in which the teacher explores the learner's learning and thinking processes (Gibbs & Stobart 2003, Lunt 1994). Calling this 'assessment' implies that, ideally, all assessment should be like this, and that teaching is ultimately the same as assessing. In SQA, we prefer to distinguish different requirements for assessment according to its purpose, for example, where assessment is used as a contribution to teaching, a tool for accountability or a requirement for certification. Which contribution assessments make to teaching depends on how they are used.

# Current issues

## Sampling

A major issue in Assessment is for Learning (AifL) is the tension between, on the one hand, the wish to reflect and inform activities in the class as directly as possible, and on the other hand a need to be able to conclude that what has been learned in the class meets the overall standards set by the wider system for official accreditation.

This issue can be seen from several perspectives. From the perspective of sampling, regular informal evidence-gathering will cover more of the content, thought processes and situations than an external assessment at the end of the year, but it is likely that informal internal evidence will diverge from the official interpretation of the objectives, because of the teacher's, learners', and other classroom-related input. This should not create a problem if a better use of external assessment tasks leads to better teaching and so to more progress within the school-defined objectives. It can also be expected that better quality of learning ('deeper learning') will allow learners to perform better on external assessments, even if the objectives are slightly different.

## Thought processes

Assessments can be used to check whether learners have acquired and can apply much-valued higher-order thinking processes. Assessment of thought processes depends heavily on the similarity between assessment tasks and learners' learning experiences.

Familiarity with the processes needed to produce an answer, with the data provided with the question, or with the situation in which the question is set, determines whether learners will be able to give an answer from memory, execute routine actions, or will have to analyse the problem, find a solution and evaluate whether this is correct. Teachers are in the best position to determine what is familiar to their learners and what is less familiar. They should be able to determine not only what has been dealt with in the class, but also how situations and examples fit in with their learners' broader experiences. If teachers were well trained in analysing subject content and thought processes and in writing assessments, and if they also had the time available to write their own questions, their assessments would be near perfect.

On-line systems that contain both teaching and assessments within the same environment are in second place when it comes to matching assessment and teaching. External assessments however, tend to suffer from repetition and over-exposure. Even questions

which required the highest level of analysis and synthesis when asked for the first time, can become routine tasks after having been cloned and practised in subsequent years.

### **Achieving standards**

It is generally known that learning improves when learners know the standards that are expected of them, ie when they know what they are learning and how well they will need to be able to demonstrate it. In Sweden, schools are obliged to inform learners about the standards, and schools which do more than just that have been shown to be successful. This practice relies on a thorough and common interpretation, among teachers, of standards and outcomes, which are usually more general than what is being taught at any given moment in class. Central, external assessments, arrangements, and support material cannot sufficiently clarify these standards for every possible local situation, because local situations are determined by a combination of context and presentation specific to learners', teachers' and schools. Other moderation or developmental activities are required, certainly if external assessments are to be replaced partly or wholly by teacher assessment (see Wiliam 2006).

## **Current types and purposes of assessments**

### **Psychological tests**

Some tests have been carefully constructed to either measure traits, such as intelligence, or concentration, or to identify specific problems, such as dyslexia, or spelling. These tests rely on standardised administration and compare a learner with a norm group. The development and standardisation of these tests is expensive and requires expert knowledge. The administration and interpretation requires training as well, and is usually left to educational psychologists.

### **Diagnostic tests**

Diagnostic tests can be focused more on subject specific concepts which have been found to pose difficulties to some learners. These tests can vary from almost psychological tests, to teacher made assessments concentrating on concepts which part of the class may not have mastered, despite having made more than one attempt. The teacher made assessment can take into account the processes taught and the material used with the learners, and to a certain extent their individual wider experiences. The quality of such teacher made tests clearly depends on the experience the teacher has in recognising problems and in analysing relevant concepts and thought processes. Where this assessment process is imperfect, it can be expected to correct itself, because it should become clear from unsuccessful next steps whether the assessment or conclusions drawn from it were invalid.

## **Monitoring progress**

As it is considered to be very important in the educational system to know how much progress is being made towards levels, aims and objectives (if not targets), assessment often concentrates on finding out how much learners have learned. This can be done informally, or formally. In the worst case, this type of assessment intensifies the focus on accountability without contributing to the quality of teaching. If there is pressure to show progress to parents, and to achieve targets, teaching is likely to imitate assessment. This is when assessment influences teaching most negatively, when assessments cover small periods of teaching, take considerable assessment time, and assessment practice replaces teaching. In this situation, assessment is not likely to take into account the situation, experience and processes shared by a particular teacher with a particular group of learners. Nevertheless, systems to monitor progress may be presented as a tool to improve teaching, suggesting that identifying gaps is practically the same as identifying reasons and remedies. At best, such systems provide a fine-grained profile based on a common curriculum, and offer on-line activities for remediation.

## **What can an examination body contribute to internal assessment for learning**

Possible contributions by SQA to assessment for learning in the class are of three types (see Neil Jones 2006):

- ◆ Development
- ◆ Ready-made assessments
- ◆ On-line systems

### **Development**

SQA can contribute to the understanding and application of its standards by further development of its online support, such as that provided on the SQA Academy and more specifically by making this accessible to learners as well as teachers.

SQA could also support quality assurance systems, which do more than moderating internal components of external exams. From the experience so far it is clear that development should aim at changing behaviour, not at spreading knowledge. Williams (2006) contrasts the low impact of knowledge-based professional development on teachers' behaviour with the considerable impact of programmes such as WeightWatchers.

## **Ready-made assessments**

External educational organisations will be better placed than teachers to develop costly and specialist assessments. Some subjects, for instance languages, are more performance and ability based than content-oriented. Developing assessments of levels of speaking for instance, is not only very complicated, it also seems to be done best by applying different statistical methods from those used traditionally to create appropriate scales (see Neil Jones 2006). The complexity arises because there is a variety of criteria and a range of unwanted influences on assessment. These assessments are best used by teachers who are trained to administer them and interpret the results. Of course, commercial organisations offer assessments as well, either linked to textbooks, or public curriculums, or attractive for other reasons.

On the face of it, the existing NABS do not seem to contribute easily to any of the five main strategies for Assessment for Learning as formulated by Williams 2006 (although they can be turned into activities which support the first two strategies):

- ◆ clarifying and understanding learning intentions and criteria for success
- ◆ engineering effective classroom discussions, questions and tasks that elicit evidence of learning
- ◆ providing feedback that moves learners forward
- ◆ activating students as instructional resources for each other
- ◆ activating students as owners of their own learning

## **On-line systems**

Because on-line systems are expensive to develop, and require a combination of other expert skills over and above those of experienced teachers, they are typically offered by examination bodies, large publishers, or educational software developers. For examination bodies this offers an opportunity to strengthen and employ their computer aided assessment expertise and facilitate a smooth transition to on-line assessment for certification purposes, as well as providing a source of intelligence on the uptake and use of non-certificated courses.

## What effect could Assessment for Learning have on assessment for certification?

Assessment for Learning places more emphasis on understanding of assessment tasks in relation to the standards, to thought processes and content (to provide feedback), and to the learners' needs. It is certainly not looking for exemplar external assessment tasks that can be copied, used, and used repeatedly in the classroom. Even external assessments for certification that were deliberately designed to be used for formative purposes as well, would be likely to be abused for exam practice instead of for teaching.

Assuming that external assessment for certification will continue to be used at SCQF level 4 and 5, it might be possible to improve the alignment with formative assessment by ensuring that the external assessments have:

- ◆ Authentic (ie real-world) tasks and situations.
- ◆ Clear explanations of what each task is meant to assess (in terms of content, skills, thought processes, and situation). A large part of these clear indications are already available, but distributed over several sources. Would it be possible to use (a modern version of) a test grid/description/matrix?
- ◆ Clear indications of the criteria for success, related to the certification (not just the boundaries, but some explanation for the usual boundaries based on content such as grade and outcome related criteria).
- ◆ Transparency about the quality of the assessment, for instance standard error of measurement in terms of grades, as well as average scores for components.
- ◆ Transparency about the limitations put upon the use of the results as a result of the quality and type of qualification. Examples are warnings not to see differences in grade as absolute and reliable differences in the underlying ability of candidates ('She is an A, he is only a B'), or generalising from a few subjects to general academic ability.

## What effect could internal Assessment for Learning have on assessment for accountability?

Were the certification and the accountability functions of existing exams to be separated, it would be an important change.<sup>1</sup> Certification requires an exam for every single learner who

---

<sup>1</sup> The new Framework agreements allow this, as they concentrate on a few general indicators for an authority as a whole. Other mechanisms are changing too. An important distinction is that between quality of teaching and amount of attainment.

opts for a certificate at the end of the course (and is usually assumed to apply to everyone following a course). Accountability requires sufficient information at several levels in the educational system, and can be operated by sampling, as in the SSA (see recent announcements on assessing the National Curriculum in England and Wales, and Sylvia Green et al (2006). Sampling allows the coverage of standards to be finer grained and more comprehensive. The SSA has managed to include types of tasks excluded from external examinations, such as oral interviews and practical tests. This shows that the accountability function can be better and more efficiently fulfilled (at national and authority level) if based on sampling according to the SSA model, than by simply aggregating assessments.

## References

Dylan Wiliam (2006). *Assessment for learning: why, what and how. Excellence in Assessment: assessment for learning*, Cambridge September 2006.

Neil Jones (2006). Assessment for Learning: the challenge for an examination board. *Excellence in Assessment: assessment for learning*, Cambridge September 2006.

Louise Hayward, Nicki Hedge (2005). Travelling towards change in assessment: policy, practice and research in education. *Assessment in Education*, 12,1 55-70.

Bertil Roos, David Hamilton (2005). Formative assessment: a cybernetic viewpoint. *Assessment in Education*, 12,1 7-20.

Dylan Wiliam, Clare Lee, Christine Harrison, Paul Black (2004). Teachers developing assessment for learning: impact on student achievement. *Assessment in Education*, 11,1, 49-65.

Caroline Gipps and Gordon Stobart (2003) Alternative Assessment. In: T Kellaghan & D L Stufflebeam (eds) *International Handbook of Educational Evaluation*.

Sylvia Green, John Bell, Tim Oates & Tom Bramley (2006). *Alternative Approaches to National Assessment (KS1, KS2, KS3)*. Internal paper Cambridge Assessment.