

Approaches to internal assessment as part of examinations for certification purposes

Paul Black, Emeritus Professor of Science Education, King's College London

1 Justification

Internal assessment involves costs in administration and in teachers'/lecturers' time, which may be fully justified if it enhances assessment validity by reflecting important educational aims which externally set and marked question paper examinations taken under controlled conditions cannot assess. Reliability may also be enhanced because learners can be assessed in more varied, less stressful, and more frequent occasions, than is feasible with externally marked question papers or internally administered performances. In addition, the time teachers/lecturers will have to devote to internal assessment work can make a positive contribution in that, by taking more responsibility for certification assessment they will be helped to develop a more productive relationship between learning and assessment.

2 A moderation/verification system

Assume that the internal assessment results are recorded as a judgement in terms of a mark (or grade or level), and evidence in relation to which that judgement can be justified and checked. The evidence may be a single artefact or a collection of several artefacts, an 'artefact' being a written document, or a constructed object, or an audio or moving image recording, or a combination of these. Assume that the assessment results are produced by the following three-stage process:

- a) A centre, through collaboration between the subject teachers/lecturers, produces a set of assessment benchmarks, agreed between them by selecting samples from across their classes and blind-marking¹ these before a meeting to compare results, agree interpretations of criteria, and resolve discrepancies.
- b) A group of centres follow the same procedure, starting with benchmarks selected by each school as a result of process a). These marked benchmarks then serve to guide all the marking.

¹ Blind marking involves giving each marker a copy of the work, or opportunity to view it, with no indication of the judgements of others in the group, and recording individual marks and comparing these at a subsequent meeting.

- c) The overall results, together with the benchmarks from the centre, are then submitted to the certification authority who may, on the basis of statistical analyses and/or scrutiny of benchmarks, either accept, or adjust the results, or call for further evidence.

For the system to be **workable**, there has to be adequate comparability between the evidence produced by different teachers/lecturers, centres, and groups of centres. For it to be **justified**, the portfolios should comprise work which is valid in relation to agreed curriculum aims. The rest of this paper explores the various stages in the production of a portfolio, in order to detail the judgements that have to be made to achieve the optimum alignment between these two types of criteria.

3 Production of a portfolio/case study/assignment/etc

This section describes a sequence of steps, each of which needs careful consideration because it influences the achievement of the two criteria.

First: Any individual tasks or artefact should be so designed that learners' responses will represent valid achievements, should allow for a full range of discrimination, giving opportunities to both low achievers and high achievers to demonstrate their capabilities, whilst also engaging and motivating learners so that all produce their best work.

Second: Tasks must be so presented to learners that they have a clear understanding of what they are expected to do, including understanding the criteria of quality which will be used to judge their work: any explanation which can affect the task demand should be the same for all learners.

Third: The conditions under which learners produce their work should at least be on record, perhaps specified: relevant features include the time allowed; whether the work is entirely within classroom conditions, or involves work outside classes; whether the resources available are specified, controlled, or left open to individuals' initiative; whether collaboration between learners is allowed, or even encouraged; and whether the conditions specified for preparation (eg collecting data) differ from those specified for production of the final piece of work.

Fourth: The portfolio/case study/assignment/etc will comprise the set of tasks on which a learner's result will be based: there could be only a single task, or two or more designed to reflect a range of types of challenge, giving learners more opportunities to achieve.

Fifth: For the way in which marks are assigned, protocols should relate directly to the unit specifications, should specify marking scales, whether in raw numbers, or percentages, or in relation to grades or levels, and must also include protocols for aggregation of marks, within tasks and/or across several tasks, to determine the final outcome.

Sixth: Procedures/rules for submission of marks and for selections of samples have to be formulated and agreed in preparation for the work of the **moderation/verification system**.

4 Achieving validity and reliability

The first four of these steps are important for securing validity of the portfolio evidence, so that the use of this evidence is **justified**. However, the marks produced can only be fair and comparable across different teachers/lecturers and schools and colleges if there is some uniformity across the different aspects which are detailed above, otherwise moderation groups will not be **workable** as they will be trying to compare the incomparable. Compromises will be needed, for very rigid rules may force all teachers/lecturers and learners to follow exactly the same procedures and so encourage repetitive use of tasks, which can undermine the desired validity. Examples of practical compromises might be:

- a) In a portfolio of six tasks, all teachers/lecturers and schools and colleges in a moderation group may use the same three, with the other three left for individual teachers/lecturers to choose or create their own.
- b) Learners might undertake collaborative preparation work to explore and use a range of resources and skills to produce posters and assess them, under flexible conditions, and then produce, working individually in test conditions, a different poster in a two-hour exercise, with each provided on the spot with a set of resource materials to select and deploy.
- c) A teacher/lecturer, on seeing that a particular learner has so misinterpreted a task that he cannot gain any credit for his work, may intervene to help that learner grasp more clearly what is required so that he can show what he is capable of, with the teacher/lecturer recording this extra help so that any marker may allow for it.

It is hard to list all possible compromises, let alone select an optimum list in terms of the cost/benefit balance for each. One reason for this is that the constraints and opportunities will depend on the nature of assessed tasks: a report on fieldwork in Geography, a piece of creative writing in English, and design and construction of an artefact in Technology may all call for different compromises. Some of the choices involved may be devolved to school or college groups, to schools or colleges or to individual teachers/lecturers, others will have to be made and imposed centrally. Such centrally determined rules may offer options, eg either use a specified task, or use your own under more tightly prescribed conditions (as in the Ofqual Controlled Conditions system).

5 Assessment and communication of and about learning

If a system based on group moderation/verification is to work well, the group may have to meet on two or three occasions during the year to develop their shared understanding, and to justify any novel or individual procedures, so that potential problems are anticipated well before the final high-stakes moderation event. Teachers/lecturers will have to bring to such meetings, and share in them, their understandings and new ideas about such issues as validity and their interpretations of the curriculum objectives. Such meetings have been shown to be valuable continuing professional development (CPD), helping to develop the participants' skills and concepts in assessment matters. An explicit and shared approach to assessment procedures can also improve communication, between teachers/lecturers, and between teachers/lecturers, learners and parents and other stakeholders. All these features can enhance the quality of summative assessments throughout a school or college.

Background references

For a detailed description of what might be involved in a fully developed assessment system with heavy reliance on assessment by teachers, see pp 18 to 34 on Standards Based Assessment in Australia see:

- ◆ Stanley, G., MacCann, R., Gardner, J., Reynolds, L. and Wild, I. (2009). *Review of teacher assessment: what works best and issues for development*. Oxford University Centre for Educational Development: Report commissioned by the QCA. Downloadable from Oxford Centre for Educational Assessment.

For accounts of the issues directly relevant to Scotland, see:

- ◆ Hayward, H., Dow, W. and Boyd, B. (2008) *Teachers' Understanding of Standards*. Report on Research Project to be published by LTScotland.
- ◆ Boyd, B. and Hayward, L. (2008) *Exploring Assessment for Accountability*. Report on Research Project to be published by LTScotland.

For an account of an exploratory project to develop the quality of teachers' coursework assessments in England, see:

- ◆ Black, P., Harrison, C., Hodgen, J., Marshall, M. and Serret, N. (2010) Validity in teachers' summative assessments. *Assessment in Education* 17(2) 215–232.
- ◆ Black, P., Harrison, C., Hodgen, J., Marshall, M. and Serret, N. (2011) Dependability of teachers' summative assessments. For publication in *Assessment in Education*.