School-based Assessment and School Autonomy

Discussion Paper 11

Abstract: School-based assessment in Queensland means that teachers have responsibility for constructing and administering assessment instruments, and for appraising student work. But because certificates are issued from a central authority, the assessments must be comparable from school to school. In addition to being school-based, the ROSBA system is criteria-based as well. It is argued in this Paper that using uniform criteria and standards across the state allows for variety of approach in assessment and helps to achieve comparability without destroying the autonomy of the school.

Author: Royce Sadler, Assessment Unit, April 1986

Note: This Discussion Paper has been produced by a member of the Assessment Unit of the Queensland Board of Senior Secondary School Studies. It is recommended to the teaching profession for consideration in the formulation of curriculum and assessment policy within secondary schools. Reactions to and comments on its contents would be helpful.

© This material is copyright. It may be copied freely for the use of schools and teachers in Queensland. It may not be reproduced for sale without express permission.

Compared with external examinations, school-based assessment delegates a number of responsibilities to schools with respect to the summative evaluation of students’ achievements. The domains and the degrees of autonomy actually accorded to the schools must be resolved under such a system. It is obvious that there cannot be absolute autonomy in any situation where there is a requirement for comparability, such as occurs when certificates are issued by a central agency (which in Queensland is the Board of Secondary School Studies). Schools therefore enjoy what might be called a ‘bounded autonomy’. It is necessary to make the boundaries clear if public confidence is to be maintained, and if schools are to operate efficiently (and feel comfortable) in the system. Where should the line be drawn between complete freedom for the school, and complete control by the central authority? In this Discussion Paper, some of these issues are explored. Comparability of exit Levels of Achievement is completely consistent with the philosophy of ROSBA (and with the spirit of Radford before it). To hold that ‘school-based’ means that schools have sovereign rights to determine exit Levels of Achievement according to criteria and standards which they set themselves is to extrapolate beyond the intentions of both the Radford and ROSBA committees. Neither of their reports implied freedom to that extent, and both outlined a set of measures to ensure that standards would be the same across the state.

At the outset, however, it is as well to point out that substantial freedoms exist whether the certification system employs external examinations, a mix of external examinations and school assessments, or school assessments alone. Under any of these systems, schools normally have certain curriculum freedoms, including the freedom to select content specializations (in some subjects), and to design appropriate learning experiences (in all subjects).

With respect to assessment, schools typically are free to decide on the
(a) nature and quality of feedback to students,
(b) total time and energy devoted to assessment during the course,
(c) frequency and timing of assessment activities,
(d) form and manner of reporting to parents,
(e) items, tests, and other instruments to be used for purposes other than final certification.

None of these can be ‘claimed’ as a benefit of school-based assessment as such. The freedoms available under school-based systems lie elsewhere. There are three distinctive features of school-based assessment, which apply whether the basis for reporting student achievement is norm-referenced or criterion-referenced. The first feature radically extends point (e) above to the extent that teachers themselves make the summative judgments about the extent of students’ knowledge and the quality of their work for purposes of certification. To do this, teachers design and administer their own data-collection programs. The second feature is a consequence of the first: schools are free to decide on the balance between formative and summative assessment. The third feature is that school-based assessment relies on a distributed professional responsibility for operating a system of checks and balances in order to achieve comparability. (In Queensland, this involves officers of the Authority and experienced teachers who sit on District and State Review Panels. The administrative structures for this system are normally under the control of the same central authority as issues the certificates.

If, in addition to being school-based, the assessment is to be criteria-based, both the assessment instruments and the teachers’ judgments must be related to the criteria and standards. A strong argument can be mounted for these criteria and standards, in the interest of comparability, to be laid down in the syllabuses. Although this latter requirement places some constraints on schools, there remains considerable latitude in the ways assessment programs can be designed and still be in accordance with specified criteria and standards. It should be obvious, however, that if a criteria-based assessment system is to realise its intentions, the principles and practice of assessment using criteria and standards must be thoroughly understood by the users.

In the operation of a criteria-based certification system, a useful analogy can be drawn with the practice of calling tenders for a product or job. The client makes available the specifications for what is required, and suppliers examine how they can fulfil the requirements. The client is typically less interested in exactly how the supplier proposes to meet the minimum requirements than that the requirements be met somehow. This difference between process (means) and product (end) is, as it turns out, also reflected in standard tendering practice. Specification writers call the former ‘descriptive specifications’ and the latter ‘performance specifications’. Where facilities are available for checking the quality of the end product, performance specifications are generally preferred because they give the supplier of goods or services the greatest possible latitude in selecting a method (or devising a new one) to fulfil the requirements.

Tendering practice provides a perspective from which to view ROSBA. The Board lays down certain specifications (criteria and standards) for the various Levels of Achievement in each subject. Each school examines its clientele, its philosophy, and its human and capital resources, and devises a scheme whereby it can meet the Board’s specifications with a certain form of ‘product’, namely, student achievements. As part of the reviewing process, a school tenders its proposal to a Review Panel, which then accepts, suggests modification to, or in rare cases rejects, the proposal as being adequate for satisfying the specifications for purposes of certification. Unlike the commercial analogy, ALL tenders that satisfy the requirements are, of course, accepted. In some subjects, the ‘products’ offered may look superficially quite different from school to school, because of the nature of the subject. In other subject, the constitutive elements of the subject, a long tradition of certain forms of assessment, or almost universal use of a particular textbook, may result in essentially similar products. Naturally, greater variety can be expected in some subjects than in others. In either case, the processes the school uses in helping its students satisfy the standards are of only incidental interest at the point of certification.

From a school’s point of view, meeting the standards at exit includes:

(a) organizing learning experiences in such a way that students are placed in a position where they can achieve to the limit of their potential;
(b) setting interim or progressive targets along the way, semester by semester, so that students have realistic short-term goals that are achievable with the resources available; and
(c) tendering, and making the case as to how the end-of-course achievements of students at the school match with the standards laid down in the syllabus, that is, making the case for ‘equivalent to (in quality) but different from (in detail)’ what is required for the award of the various Levels of Achievement.

The logistics of getting such a system going in the first place are, as it turns out, quite complicated. The criteria and standards contained in the syllabuses apply to the work of students at the end of the two-year courses, but the schools have to receive approval (or accreditation) on the basis of a proposal for assessment submitted at the beginning of the course. At that point, schools have no concrete examples of student work to display, only prospective achievements. Once the initial hurdles are overcome, however, both schools and the Board can refine the processes and develop confidence in using criteria-based assessment.

**Stimulating variety**

A central authority has great potential for facilitating change, especially when it not only advocates that distinctiveness among schools is educationally desirable, but also shows how it can be accomplished. Ordinarily it would be insufficient merely to provide the conditions in which differentiation would be possible if at the same time schools were left feeling somewhat insecure about what would be acceptable. Ironic though it may appear at first, a central agency by adopting proper strategies can actively promote change and variety, and still fulfil its obligations to society for comparability. The essence of the ROSBA system is that the responsibility for the specifications as to quality of the end product remains with the Board, but that schools formulate their own methods of meeting the specifications.

A difficulty at present is that many teachers lack the confidence or the skill to prepare ‘equivalent to but different from’ tenders. One way to attack this problem would be for the Subject Advisory Committees to formulate clear criteria and standards, and then to compile (say four or five) different sets of exemplars which are accepted as legitimate ‘realizations’ of the standards on the criteria. In other words, the Subject Advisory Committees could take on board the task of assembling and disseminating multiple exemplar sets that indicate not only the scope of variation possible, but also make it easier for the teaching profession and the public to appreciate the validity of different approaches. An important practical spin-off could well be that the criteria and standards come alive in the presence of difference exemplars. Otherwise there could be a tendency for them to remain as abstractions, on which teachers could all say they agree, while understanding them differently!

It is argued in the next section that exemplar sets would be meaningless without details of the associated criteria and standards to hold them all together, and that the use of statewide criteria does not have to result in tight prescription and bland uniformity. The combination of clear criteria and standards, accompanied by examples of diversity, could do more to encourage heterogeneity in syllabus implementation, local (school-level) specialization in approach, and experimentation with novel forms of assessment than exhortations for schools to branch out. At the same time, it could help fulfil the conditions for equivalent standards across the state.

**The mediating role of criteria**

The ROSBA system places great value on both variety and comparability, factors which may on the surface appear to pull in opposite directions. In this section, it is shown how the appropriate resolution lies not in some sort of pragmatic compromise, but in a proper understanding of the crucial role that can be played by carefully formulated criteria. The argument itself is an indirect one, and begins with the concept of comparability. What does it mean to say that things are comparable? Actually, there are two easily distinguished meanings recognised by compilers of dictionaries.

The first meaning, which is “able to be compared” with the emphasis on the able to, follows directly from the etymology. It is used when there is some question as to whether it is possible to compare two things
that are not identical, and is written here as compare-able. The second dictionary meaning is “more or less equal”, and when used this way, the word is written here as com-parable. The proposition about to be argued is that compare-ability of two potentially equivalent things is logically prior to a judgment about the equivalence itself. (In the present context, the “things” are student achievements from different schools.) Compare-ability is essentially prospective, and com-parability descriptive, of the drawing of a conclusion as to similarity or equivalence. Criteria, it is argued, play a decisive role in making such judgments.

First consider compare-ability. Some things so obviously belong to the same genre, and are so obviously compare-able, that the matter is not given a second thought. For example, performances of a set piece of music by different students are clearly compare-able. So are the acrylic bowls made by students in the manual arts workshop. The relative merits of things from within a particular genre can often be decided without resort to explicit criteria. In other words, the criteria are implied by the context, although they have to be brought to the surface and made explicit if teachers’ qualitative judgments are to be substantiated to students, other teachers, and parents.

The question of compare-ability for things from different genres (such as a watercolour and an oil painting in art, or a cartoon and a poster in English) is more complicated because the things may be compare-able with respect to some features or aspects and not with respect to others. For example, an egg and a glass of milk are not compare-able with respect to shape (compare-ability in this case is a nonsense idea), but they are compare-able with respect to nutritional value of cholesterol content. Here, a criterion by which to assess compare-ability is not immediately suggested by the context and has to be stated explicitly. Once it is stated, the compare-ability issue is settled and it becomes sensible to ask whether the things might also be com-parable.

The need for criteria which cut across several (but not necessarily all) genres is directly dependent on the degree to which surface dissimilarities may obscure the deeper commonalities. But it is precisely these commonalities (which are signalled by the criteria and standards in a subject) that help to characterise art as art, and English as English. Under school-based assessment, schools make choices with respect to such aspects as content and assessment instrumentation, within the general framework of a syllabus. As a consequence, subjects are ‘expressed’ differently in different schools. Uniform criteria (and standards), by providing common links between these different expressions, make compare-ability possible.

There are, of course, things that belong to genres that are apparently so distant from one another that no common aspect springs to mind at all. For example, most people would say that a banana and a truck are simply not compare-able (although even here size is a possible criterion). The ROSBA system can be thought of as pitched at an intermediate level: neither complete uniformity, nor bananas-and-trucks, but different expressions of the same ‘essence’ of a subject, in which a common set of criteria and standards can be instrumental in moving towards first compare-ability, and then com-parability. That is why it is so important under ROSBA not only to have criteria, but for all school to work to a common set of criteria which, therefore, must be stated in syllabuses.

Conclusion

Under ROSBA, schools enjoy certain freedoms as to how they assess their students, but they must conform to the requirement that the tasks set for students produce outputs that are compare-able with the outputs from other schools. This means that a school cannot be free to decide its own criteria, but must ensure that the criteria contained in a syllabus can be applied to the achievements of its students in that subject. Furthermore, in the interests of accountability and public confidence, work associated with each of the exit Levels of Achievement must be com-parable in the descriptive sense. That is, the standard of work for a particular Level of Achievement in a subject has to be, within the limits of human judgment, the same from school to school.
Author/s:
Royce, Sadler D

Title:
Discussion Paper 11. School-based Assessment and School Autonomy

Date:
1986

Persistent Link:
http://hdl.handle.net/11343/115865