

**Assessment systems for the future: the place of assessment by
teachers (ASF)**

Report of Seminar 3, held in Cambridge, June 21st and 22nd 2004

Contents:

Summary

1. Introduction
2. Overview of the programme and outcomes
 - 2.1 Day 1 Presentations and discussion
 - 2.2 Day 2 Identification of issues and discussion
3. Key points from discussion of emerging issues
Outcomes from discussion groups

Appendix A: List of participants

Appendix B: Seminar programme

Summary

The third invited seminar of the Assessment Systems for the Future (ASF) project was held in Cambridge, June 21st and 22nd 2004. The overall aim of the project, which is funded by the Nuffield Foundation, is to study the practices and issues relating to the role that assessment by teachers can take for summative purposes. The purpose of this third seminar was to share and discuss experience of using assessment by teachers from countries outside the UK. The seminar was spread over two days. The programme and list of participants are given in Appendices A and B of the full report, which is on the ARG website (www.assessment-reform-group.org).

On the first day there were presentations from Professor Graham Maxwell, of Queensland Studies Authority and from Professor Mark Wilson, of the University of California at Berkeley, with respondents' comments and group discussions. Graham had provided background papers that were circulated beforehand, giving an account of the procedures for conducting and moderating teachers' assessments used in the Queensland Senior Certificate. In discussion, the key points raised concerned: the participation of a large proportion of teachers in moderation panels; the moderation procedures including the approval of each school's work programme; the assessment being continuous and so serving both formative and summative purposes; the system not being linked to school accountability. After 32 years it was agreed that the system 'works' although there was not the hard evidence of this such as might be required by outsiders.

Mark Wilson presented some theoretical discussion of the elements in any assessment as background to describing the Berkeley Evaluation and Assessment Research (BEAR) assessment system. In this system student achievement is judged against typical behaviours indicating achievement within a hierarchy of levels and provides both formative and summative assessment data. In discussion the interaction among curriculum, pedagogy and assessment was underlined, one implication being that learning activities should be as carefully developed and trialled as assessment items.

The second day began with three participants briefly reflecting upon lessons to be learned from experience in other countries. The main issues raised concerned: how to take cultural differences into account but still learn about what may inhibit or facilitate changes in particular contexts; the value of the student perspective and the importance of findings ways in which an assessment system can provide for students and teachers to share and reflect upon evidence; the need to increase public and teacher confidence in summative assessment by teachers. From these inputs and points added by other participants, a list of questions for group discussion was created. The discussions gave most attention to:

- how to ensure that assessment has a positive impact in classrooms
- the essential differences between formative assessment and summative assessment
- how to begin to build teachers' confidence and expertise in assessment
- where would be the best place to start in changing a system
- the conditions and processes that would foster greater reliance on assessment by teachers.

1. Introduction

This was the third seminar in the series of five seminars and two consultation conferences conducted by the Assessment Systems of the Future (ASF) project, funded by the Nuffield Foundation. The overall aim of the project is to study the nature, practice, potential and challenges of using assessment by teachers in summative assessment systems reported in full on the Assessment Reform Group (ARG) website. In seminar 1 we: a) explored current understandings of assessment by teachers for summative purposes and the role of teachers in formative and summative assessment; b) considered the research evidence on the reliability and validity of assessment by teacher used for summative purposes

In seminar 2 we a) discussed inputs on how assessment by teachers features in current practice and plans in the assessment systems of the four countries of the UK; b) worked towards the clarification of the terms used in describing the process and components of assessment by teachers used for different purposes.

The aim of the third conference was to explore what can be learned from the part that assessment by teachers plays in assessment systems in other countries. Invitations to present accounts of practice were accepted by Prof. Graham Maxwell, of the Queensland Studies Authority, and Prof Mark Wilson, of University of California, Berkeley. We had also planned to have an input from the OECD, with information from the case studies of different countries, but unfortunately the speaker had to withdraw at the last moment. The list of participants can be found in Appendix A.

2. Overview of the seminar programme

2.1 Day 1 Presentations and discussion

The programme is given in Appendix B. After a brief introduction to the project by Wynne Harlen, Gordon Stobart introduced Graham Maxwell from the Queensland Studies Authority. Graham had provided as background three papers, which had been circulated to participants before the seminar:

Assessment in Australian Schools: current practice and trends¹

Progressive assessment for learning and certification: some lessons from school-based assessment in Queensland²

Moderation of Teacher Judgments in Student Assessment²

In his presentation Graham reported that the assessment requirements in Queensland were

- Year 2 Diagnostic Net
- Years 3, 5 & 7 literacy and numeracy tests, with the possibility of year 9 tests
- Senior Certificate (end of Year 12)

All were based on teachers' judgements, with the literacy and numeracy assessments being supported by the use of externally produced tasks focused on the Queensland Senior Certificate (assessed in Year 12 at the end of a two year course) where there was 32 years of experience of using assessment by teachers (which he referred to as school-based assessment) for this high stakes certificate. Although changes are being

¹ Cumming, J. and Maxwell, G. S. (2004) Assessment in Australian schools: current practice and trends. *Assessment in Education*, Vol 11, No 1, March 2004, 89 –108.

² Available to download from the ARG website

made in the process of certification the basic principles of the assessment remain the same, viz that assessment is:

- continuous and formative
- based on a portfolio of student work
- matched to the school's work programme
- judged against standards at five levels of achievement

The final outcome was not a summation across the portfolio but an interpretation of the evidence, in which later work may supersede earlier work relating to the same goal.

External moderation procedures were described as involving approval of each school's work programme, monitoring of the assessment process at the end of year 11 (when five portfolios are reviewed), verification of processes towards the end of year 12 (to ensure adequate procedures are in place in the school) and confirmation of the final outcome for students. A random sample of portfolios and results is scrutinised after completion of the examination, targeting those subjects likely to have problems. External moderation is carried out by panels of moderators drawn from the schools and trained by QSA.

Of particular relevance were the points made about continuous assessment based on teachers' judgements. Continuous school-based assessment:

- Synthesises formative and summative (continual feedback on progress against outcome targets)
- Builds and updates an evidential record of progress (dispersed rather than peak pressure; no surprises at end)
- Fits new understandings of learning (context, engagement, authenticity)
- Extends range of assessable learning (broadening the curriculum)
- Develops student self-assessment and knowledge

Finally, Graham noted the advantages of moderation as encouraging teacher professionalism, empowering teachers and students and constituting 'powerful professional development'. He confirmed that moderated assessment by teachers can produce results of acceptable dependability and noted that the cost was less than for an external examination system. He advised making a start with lower stakes assessment to build confidence in teachers' judgments and accepting that it will take time to build this confidence.

John Gardner responded with some observations about the contribution of school-based assessment in all high stakes assessment and in all reporting. He also noted that there were many hidden costs in school-based assessment. This led into general discussion.

In reply to a point raised about the 'stakes' for the school, Graham replied that in Queensland, school accountability is not linked to certification. A related question as to the political environment (is it only a professional dialogue; what about the media?) was answered by agreeing that it is necessary on the one hand to educate politicians and on the other to help schools use the data in school improvement and to avoid league tables.

The question as to what evidence exists that the system ‘works’ was a difficult one to answer, without clarity about what is meant by ‘working’. Judy Sebba reported, from her own observations in two Queensland schools, that there was a far greater level of group-work than she had observed in schools in systems with external tests. Increasing teacher competence in assessment and using it for a wide range of student outcomes seems to be an important way of increasing confidence in teachers’ judgments.

Asked about the diversity in practice in assessment, Graham indicated a wide range, which, in the early years of implementation, had included a greater use of end of term examinations than recently, when there has been greater use of group work and assignments. He emphasised that common goals and criteria do not mean either common learning experiences or assessment tasks; the specific content does not matter as long as the same criteria for assessing learning are applied.

In the second presentation Mark Wilson described the Berkeley Evaluation and Assessment Research (BEAR) assessment system. He prefaced this account with comments on how current assessment systems fail to provide useful feedback to students, or teachers or administrators. He drew on the structure of three interconnected elements which every assessment should have and which are described in Pellegrino et al³:

- cognition (a model of student cognition and learning in the field)
- observations, (well-designed and tested assessment questions and tasks)
- interpretation (ways of making inferences about student competence)

Mark discussed the four principles on which the BEAR system is based:

- A developmental perspective of student learning
- A match between instruction and assessment
- The generating of quality evidence
- Management by teachers to allow appropriate feedback, feed-forward and follow-up.

In addition he stressed that the system was coherent with formative assessment and continuous; not something conducted once a year.

The example described was of assessment in higher education chemistry. The ‘big ideas’ to be understood were set out in a hierarchy of levels of sophistication so that success could be identified at one of 12 levels (four levels defined with three degrees of success within each). Examples of what behaviours students operating at these levels would typically show, and of questions or items that would assess these levels, were set out in the BEAR framework for each concept and skill. The tendency was for the items to be multiple-choice, focused on discriminating among the levels. Use of the system involves moderation sessions, involving university teachers, teaching assistants, and students in discussing decisions about judging students work in relation to the scoring levels. In some circumstances students score their work and in doing so recognise where they are and what are their next steps. Thus the assessment can serve a formative purpose.

³ Pellegrino, J. W., Chudowsky, N., and Glaser, R. (Eds) (2001) *Knowing What Students Know*. Washington DC: National Academy Press

In his response to the presentation, Paul Black emphasised the interaction between the curriculum, pedagogy and assessment. Expertise in the subject matter is needed in providing help for teachers in assessment that can be useful both formatively and summatively. The type of assessment represented by BEAR raises the question of where the differences between formative assessment and summative assessment lie (have we just created them by using assessment differently?). Teachers may not realise the limitations of their own assessment which depend upon the quality of observations that are made and thus on the underlying materials in the curriculum.

The subsequent discussion noted the different setting of education in the US where particularly in HE, but at high school also, a teacher taught large numbers of students and needed easily marked items (ie multiple choice) in order to operate a system such as BEAR. Mark pointed out that assessment needs to be brought into curriculum development in order to drive a cycle in which pedagogy is affected positively by assessment. Learning activities should be as carefully worked out as assessment items. Although the example was concerned with cognitive understanding (in chemistry) the same approach can be used in relation to processes, with observations based on practical activities.

2.2 Day 2 Identification of issues and discussion

On the second day, attention was turned to applying lessons learned from the experience in other countries to the problems of making change in the assessment systems within the UK. Three speakers began the morning by briefly reflecting on selected issues arising in the previous day's transactions. The emerging issues were collected, forming the agenda of questions for group discussions.

Judy Sebba talked about cultural differences that mean that what works in one context may not work in another. Therefore it is important to identify the differences and to manage change in assessment in the light of what may inhibit change in particular cases. She considered that in summative assessment in England we have come a long way toward assessing what students can do rather than what they can't do. It may not be helpful, however, to develop formative assessment and summative assessment separately, so there was an issue about how these can complement each other. Another concern was to make summative assessment genuinely useful at points of transition.

Carolyn Hutchinson drew attention to the importance of the student perspective. Summative assessment is a passport to the next year or stage of education and so must be able to capture what is widely accepted as important in learning. It needs to be fair, accurate and valued by all concerned. Students will value the 'passport' if it serves its purpose, which means that there must be public confidence that it is fair, valid and reliable and transparently so. She proposed abandoning the distinction between formative and summative assessment and focusing instead on effective assessment. There was an issue about how a system can provide for students and teachers to share and reflect upon evidence. She also wanted to know the key features that are needed to ensure that assessment has a positive impact in the classroom.

Gordon Stobart took up the theme of what conditions are needed to increase confidence in summative assessment by teachers. The question was to find the right

starting point and to be ready when there was an opportunity for implementation. There was a considerable effort needed, particularly in England, to recreate trust in teachers. At the same time, teachers needed some technical expertise, which is largely missing and has to be provided. We must avoid precipitate action with too high a profile and inadequate professional development, which would be likely to fail and put back any further attempt for several years.

During plenary discussion the following points were made:

- John Dunford reported that the Secondary Heads Association was proposing the creation of a group of accredited teachers as a way of developing confidence in teachers' assessments. This differed from the Daugherty proposals for Wales⁴ which recommended accreditation of schools.
- David Bartlett reported that the Key Stage 1 national trials (in 25% of primary schools) was indicating that teachers welcome the shift back to a focus on assessment by teachers but that change needs time to reverse the loss of confidence by teachers themselves in their own judgements. We should avoid the too-early evaluation of change (as in the introduction of the Foundation Stage assessment by teachers in England).
- In Wales, where KS1 tests had been removed, teachers felt liberated but the assessment by teachers is unmoderated and an HMI report indicates variable quality.
- John Bangs pointed out that the impact of the tests (eg in damaging teachers' sense of creativity) had been greater in primary schools than in secondary, where a steady state had been reached.

3.Key points from discussion of emerging issues

The key questions raised in the first session of Day 2 were:

1. What approaches are most to ensure that assessment has a positive impact in classrooms?
2. Are there intrinsic theoretical differences between formative assessment and summative assessment other than created by use?
3. What is the role of summative assessment at points of transitions, especially from school to school or school to work or other institutions?
4. How can a system provide opportunities for students and teachers to share and reflect on evidence for assessment?
5. What are the most effective steps that can be taken to build teachers' confidence and expertise in assessment?
6. Where is the best place to make a start in changing a system?
7. What conditions foster greater reliance on teachers' assessment?
8. What can be said about costs and benefits of using teachers' assessment compared with tests and examinations?

Discussion groups selected their own foci for attention, thus not all question were equally addressed. Graham Maxwell and Mark Wilson were also asked to respond to

⁴ *Learning Pathways Through Statutory Assessment: Key Stages 2 and 3*. Daugherty Assessment Review Group Final Report. www.learning.wales.gov.uk

these issues and further points were made by Caroline Gipps, Richard Daugherty and other participants in the final session. The main points arising from all these sources are brought together in the following, grouped roughly in relation to the question addressed, although there is considerable overlap across questions.

Question 1

Evidence is needed that using assessment by teachers works. Within subjects it is clear that a wider range of evidence is needed than conventional tests and examinations are capable of providing. The public needs to be convinced that change is needed and that it will bring better learning.

The close relationship of assessment to the curriculum is key to ensuring positive impact in classrooms. Assessment should be seen as a process embedded within learning contexts, not as an event separated from them. The teacher has the ability to build up a picture over a period of time.

There needs to be a broadening of the curriculum to reflect what we really want students to learn at different points of their education and emphasis on individual learning journeys. The increase of group work, called for in the Tomlinson review for instance, required teachers' assessment.

A further point (see also question 5) in ensuring positive impact is the transfer of some responsibility for learning and assessment to the student. Greater use can be made of ICT in student self- assessment.

Question 2

The procedures for formative assessment and summative assessment are the same, but the timing, context and uses differ. Formative is fine-grained, concerned with the moment to moment. Summative is coarse grained, concerned with achievement over time. An implication is that if teachers know how to do assessment effectively and efficiently, then their judgments should serve both purposes. For both the important judgement is how a piece of work relates to the overall goal.

The view was expressed that formative assessment should drive summative assessment – the latter being a 'sample' of the former. 'Sample' in this context meaning that the information sources have the same profile so that the particular sample does not make a difference.

Question 3

Good starting points would be transition points, such as from key stage 2 to key stage 3. A central element might be personal learning planning by students (as in Scotland).

Question 4

We can learn from current experience in vocational qualifications for post-16 year olds that have teacher assessment, centre accreditation, internal and external moderation etc, in both schools and colleges. It is not that there is necessarily 'good practice' in this part of the system but there is practice that can be evaluated to consider the implications of using assessment by teachers for summative purposes.

Question 5

Public confidence is as likely to be as important to politicians as it is to 'the public'. Moderation is the 'magic bullet' for increasing confidence; it not only ensures fairness across the system (which both public and politicians require) but is also proven to be a strong form of professional development.

Moving towards a system in which teachers are accredited as assessors may well reinforce public confidence. Teacher accreditation is not enough and needs to be combined with accrediting school systems and with moderation. It was possible that individual teacher accreditation would not bring the same professional development benefits as school moderation. There was a danger of a two-tier system, in which accredited teachers were privileged. Moderation procedures should avoid this. It was useful to distinguish between school-level moderation procedures and those conducted by individual teachers.

Question 6

Before making suggestions for structural change in assessment there must be technical clarity about how to operate a system using assessment by teachers. A national system based on moderated teachers' assessment is multi-layered and has inevitable weak links. What teachers do is public, so the pressure on them is greater than in the case of external examinations.

There was little agreement on suitable starting points for making changes in the system in England. One suggestion was to start at each end of the age range (foundation and 16 – 19) in the hope that good practice will spread to age groups between. Contrasting views were expressed about GCSE as a starting point for tackling the system. It was also thought that the Foundation stage, already based on teachers' assessment, should be left alone. Key stage 3 might be a better target.

Question 7

No change towards greater use of assessment by teachers can take place without a consistent strategy for professional development. The involvement of teachers was central. It was noted that in Queensland, 10% of teachers were involved in moderation panels. However it is essential to take cultural differences into account in citing experience in other countries.

Sustained knowledge of the subject is a pre-requisite for good assessment.

Since teachers will be in the public eye in this change it is essential for their voices to be heard. Not all teachers welcome greater involvement in assessment; not all have the necessary competence. There is a good deal to be done in convincing teachers as well as in providing the professional development to give teachers confidence in their own and each others' assessment. They also need help in progressing students' learning in a way that responds to assessment formatively and recognises what has been achieved summatively.

If assessment by teachers for summative purposes is to be developed, and used more extensively, the developments should be trialled in primary, secondary and post-school education over a period of a minimum of two years. Professional development that involves and empowers teachers is needed in support. There has to be some

evident compensation for teachers taking on extra responsibility: something has to go; teachers have to work 'smarter'. The involvement of students in self-evaluation has particular value in changing the character of teachers' work. Teachers' own self-evaluation is another avenue to explore.

Timing is of the essence. To move too soon would expose inevitable weak practice, which would be seized upon and likely to lead to withdrawal of support. Thus it is essential for teachers to be involved in making decisions about the shape of a new system and about how quickly change can be made.

Question 8

Costs need to be balanced by benefits. The costs of bring a group of teachers together have to be seen as supporting not just the quality of assessment but the improvement of understanding of learning goals and of ways of achieving them.

June 21st and 22nd, New Hall, Cambridge

PARTICIPANT LIST

Mr. John Anderson	Education Technology Strategy Coordinator Northern Ireland
Mr John Bangs	NUT
Mr David Bartlett	Co-ordinator for assessment, Birmingham Education Authority
Prof Paul Black	King's College, London and ARG
Ms Jackie Burnett	QCA
Dr. Joy Cummings	Griffith University, Queensland, Australia
Prof Richard Daugherty	University of Wales, Aberystwyth and ARG
Mr John Dunford	General Secretary, Secondary Heads Association
Dr Kathryn Ecclestone	University of Exeter and ARG
Ms Janet English	Head teacher, Kingsway Infant School
Prof John Gardner	Queen's University, Belfast and ARG
Prof Caroline Gipps	Kingston University
Prof Wynne Harlen	Bristol and Cambridge Universities and ARG
Ms Carolyn Hutchinson	Head of Assessment Branch, Scottish Executive Education Committee
Dr. Mary James	University of Cambridge and ARG
Ms Bertha Mc Dougall	Principal Officer for ICT based assessment and development, CCEA
Ms Caroline Macready	Head of School Performance & Accountability Division, DfES
Prof. Graham Maxwell	Queensland Studies Authority, Australia
Dr Catrin Roberts	Assistant Director, Nuffield Foundation
Dr. Gill Robinson	Head of Qualifications, Assessment and Curriculum Division, Scottish Executive Education Department
Mr Jon Ryder	Teacher, Lord Williams's School
Prof Judy Sebba	University of Sussex and ARG
Dr Gordon Stobart	University of London and ARG
Ms Anne Whipp	ACCAC
Prof Mark Wilson	University of California, Berkeley, USA

Assessment Systems for the Future

**Expert Seminar on International Perspectives on the use of
Assessment by Teachers for Summative Purposes**

June 21st and 22nd, 2004, New Hall, Cambridge

Programme

Monday, June 21st 2004

12.30 – 1.30 pm

1.30 – 1.45pm

Plenary

Wynne Harlen, Project Director

1.45 – 3.45 pm

Presentation 45min

Groups 45 min

Plenary 30 min

**Using assessment by teachers for summative purposes: the
Australian Experience**

Chair: Gordon Stobart

Speaker: **Graham Maxwell**

Respondent: John Gardner

3.45 – 4.15 pm

Tea

4.15 – 6.15 pm

Presentation 45min

Groups 45 min

Plenary 30 min

**Using assessment by teachers for summative purposes: the
American Experience**

Chair: Kathryn Ecclestone

Speaker: **Mark Wilson**

Respondent: Paul Black

7.15 pm

Dinner

Tuesday, June 22nd 2004

9.00 – 11.00 am

Presentations 30min

Plenary 30 min

Groups 45 min

Feedback 15 mins

Issues raised

Chair: Mary James

Speakers: Judy Sebba, Gordon Stobart, Carolyn Hutchinson

11.00 – 11.30 am

Coffee

11.30 am– 12.15 pm

Chair: Wynne Harlen

Speakers: **Graham Maxwell and Mark Wilson**

12.15 – 12.30 pm

Final comments

Speaker: Richard Daugherty

12.45 pm

Lunch and depart