

Report of ASF Seminar 4, held at Fitzwilliam College, Cambridge, January 11/12 2005

Summary

1. Introduction
2. Discussion of issues arising from consultation conferences
 - 2.1 Issues relating to accountability
 - 2.2 The curriculum/assessment interface
 - 2.3 The relationship between formative and summative assessment
3. Review of project plans and outcomes
 - 3.1 Goals and priorities for ASF
 - 3.2 Further project action
 - 3.3 Main points in reports from groups
 - 3.4 Further development

Appendix A: List of participants

Appendix B: Seminar programme

Appendix C: David Bartlett's presentation on summative assessment and accountability

Appendix D: What is an AifL School? (Carolyn Hutchinson)

Appendix E: Formative and summative assessment – a harmonious relationship? (Wynne Harlen)

Appendix F: Wynne Harlen's presentation on formative/summative interaction

Appendix G: Considerations in the design of summative assessment systems which incorporate teacher-led assessment (QCA paper by Paul Newton)

Appendix H: Can we raise the level of debate on teacher assessment? with annexes 1, 2 and 3 (Paul Newton)

Summary

The fourth ASF seminar was held in Cambridge, January 11/12 2005. Participants were members of the project Core Group plus one invited visitor as an observer. The aims of the seminar were to review the first year's work, particularly the outcomes of the three consultation conferences held in November and December, 2004, and to consider the plans for the final year of the project. Three of the issues most frequently raised in the conferences were chosen as the subjects of sessions on the first day, each being led by a member of the Core Group. These topics were: accountability; the curriculum/assessment interface; and the relationship between formative and summative assessment. On the second day, the focus turned to further work of the project and how it might identify and evaluate preferred systems of summative assessment.

1. Introduction

The fourth 24-hour seminar to be held by the project was attended by members of the Core Group plus one overseas visitor, as an observer. (See Appendix A for the list of participants). Originally it was the fifth seminar that was intended to be restricted to the Core Group members, with the fourth being one where ‘users’ views would be sought. However, the programmes for the fourth and fifth seminars were interchanged to allow the group to reflect on the three consultation meetings that took place in November and December. Hold three such meetings, with different constituent groups, instead of the initially planned one conference, had drawn attention to issues that needed to be thought through in considering the future direction of the project’s work. The programme (see Appendix B) was designed to allow general reflection on the outcomes of the conferences followed by inputs from various members of the Group on three main identified issues. These were focused on accountability, the interaction of assessment and the curriculum and the relationship between assessment evidence for formative and summative purposes. On the second day, analytical considerations of the nature of assessment systems and of assessment by teachers led into group discussions of possible ways forward for the project.

2. Discussion of issues arising from consultation conferences

Reports had been circulated from the three consultation conferences: one held in London for policy-makers from England; one held in London for practitioners in England and their organisations; and one held in Glasgow for policy-makers from Scotland, Wales, Northern Ireland and Ireland. A brief report with programme and lists of participants can be found under Note of Consultation Conferences.

Following the conferences, the project director had produced a further draft of Working Paper 1 (draft 10) including a draft of Part 3 in which revised recommendations were discussed. These changes incorporated the key points from the conferences, but did not fundamentally alter the direction of the project to date. It was agreed that this should be put on the ARG website (ASF section) as a draft indicating the outcome of the first year’s work.

The key issue arising from the conferences were those concerned with accountability, the interaction of assessment and the curriculum and the relationship between assessment evidence for formative and summative purposes. These were considered in separate sessions of the seminar, each being introduced by a member of the Core Group.

2.1 *Issues relating to accountability*

David Bartlett introduced this discussion, making clear that his perspective was that of an advisor in England. He spoke to some slides (Appendix C) beginning by considering the users of information about the effectiveness of teachers and schools. He then considered the questions that are asked about teachers and schools by different users of information for different purposes. This indicated the information needed for school review and monitoring, for in-school records and target setting and for transition of pupils from school to school. The currently used evidence was then set out and compared with evidence necessary if teachers’ assessment were to play a

more significant role. This led to the identification of series of issues that had to be addressed if change was to take place.

An overarching issue was the lack of trust of teachers that underpins the accountability structure in England. David considered it unlikely that the use of data for accountability would change – that is pupil data would continue to be used as key evidence. However, the source and range of data could be changed. For instance, the PANDA (performance and assessment record) data could be extended and moderation introduced to make teachers’ judgements more acceptable. Moderation across schools and key stages would not only provide better judgements and professional development but also more collegial relationships among teachers and schools, helping to establish and maintain trust. There was some discussion as to the extent to which collegiality could replace competition. However it was pointed out that competition was not unhealthy providing the impact was genuine improvement and not short cuts to the appearance of success (eg by teaching to tests). Currently a culture of fear of failure had been created through blaming rather than supporting lower achieving teachers and schools. Change, David pointed out, had to take account of other issues, such as workload, funding, timescales and the need to develop understanding of how various kinds of data can be used.

2.2 *The curriculum/assessment interface*

This discussion was introduced by Carolyn Hutchinson and Anne Whipp. Carolyn circulated and spoke to sections from various documents relating to assessment and the curriculum in Scotland:

- Pages 5, 8, 9, 10 from Assessment, Testing and Reporting 3-14 (Scottish Executive)
- Pages 14, 15, 16,17, from Ambitious Excellent Schools our Agenda for Action (Scottish Executive)
- 12, Purposes framework for the curriculum from the Curriculum of Excellence Review Group
- General background on curriculum review priorities from the Ministerial Response to the Curriculum review Group
- Recognising Achievement S1-S3. (A proposal for looking at assessment in S1-S3 relating to the associated curriculum review. Draft, in confidence – still subject to revision, but one way of doing it consistent with the Scottish approach to managing change.)
- What is an AifL School? (Part of initial thinking on what being an ‘AifL school’ might mean and will eventually form the basis of guidance to school managers.)

Carolyn emphasised that ‘Assessment is for Learning’ involved links between curriculum, learning and teaching and assessment as three points of a triangle. These were linked in pairs as assessment AS learning, assessment OF learning and assessment FOR learning. She invited participants to identify the key features of ‘Of’ and ‘for’ learning on the diagram. (see Appendix D).

Anne tabled a paper raising questions emanating from considering what ‘a full range of students’ attainments’ might mean, how they might be assessment, what evidence would need to be collected, what criteria might be used, how progression was to be identified. The clear implications for curriculum planning and development were that

these questions about learning outcomes (how wide a range – at what stages?) needed to be addressed before their assessment could be tackled. In many aspects of aims such as ‘pursuing a healthy and active lifestyle’ a theory of progression was missing.

In discussion it was pointed out that pedagogy has to be brought into the picture. Older models of teaching, particularly of transmission of knowledge from a teacher, are not suited to achieving some more modern outcomes. There needs to be emphasis on new ways of teaching to reach the kind of understanding and skills that are needed by future citizens. In relation to assessment, not only was it necessary to collect evidence of new outcomes but also to process the evidence differently, bring assessment more closely into teaching. The notion of assessment as learning captures this by recognising the role of students in self-assessment and setting and monitoring their own goals.

2.3 *The relationship between formative and summative assessment*

Wynne Harlen spoke to a pre-circulated paper (Appendix E) using the presentation in Appendix F and Gordon Stobart responded. The extent to which evidence collected for formative use could also be used for summative assessment, and evidence collected for summative use could be used for formative assessment was discussed. (The word ‘evidence’ was preferred to ‘information’.) The arguments led to the suggestion that each type of evidence could contribute to both uses but summative evidence could not be sufficient for formative purposes and formative evidence would need to be re-interpreted and quality assured for summative purposes. Wynne also suggested that there may be a range of purposes between formative and summative; eg informal formative, formal formative, informal summative and formal summative. She elaborated Figure 1 in the paper by including lesson goals as a focus of assessment for learning, with developmental criteria being used in relation to evidence from a range of related activities.

For some members this was making too much of the difference and it would be better to think in terms of ‘good assessment evidence’ rather than evidence specifically *for* learning or *of* learning. However, there was a danger that the criteria for ‘good assessment’ would favour the characteristics of summative assessment.

3. Review of project plans and outcomes

3.1 *Goals and priorities for ASF*

This session was led by Paul Newton and Paul Black. Paul N had circulated two papers. One was a QCA paper entitled *Considerations in the design of summative assessment systems which incorporate teacher-led assessment*. (Appendix G) The other *Can we raise the level of debate on teacher assessment?* with three Annexes (Appendix H). Paul N drew attention to the work already on-going in using assessment by teachers for summative assessment. In analysing the meaning of ‘good’ assessment by teachers, he distinguished between ‘good’ in supporting valid inferences and actions (ie giving valid information about pupils) and ‘good’ in that the process has positive educational impacts in relation to the overall assessment system. An assessment may not necessarily be good in both these ways and there may need to

be a trade-off between them and one influenced by resources. There was no ‘holy grail’ that gave highly dependable assessment results with a full range of educational benefits. This led to the argument that it was essential to prioritise in designing an assessment system.

Paul N also distinguished between a system and a ‘model’ (also referred to in discussion as a sub-system). He considered factors affecting the operation of a ‘pure teacher assessment model’. Using results of teacher assessment was open to threats from using the information for accountability purposes. He also considered that the assessment load on teachers would interfere with successful implementation. However, he considered that a hybrid system integrating tests and teacher assessment carries the risk of ‘extracting the worst from both worlds.’ Just adding some teacher assessment to a test-based system would not help and would increase the resources required. In his second paper Paul argued that the current level of debate of the issue is too low to support effective decision-making. He suggested that the ASF project could identify and communicate the pros and cons of certain systems involving teachers’ assessment. In particular he suggested that we should present the arguments for alternative systems, not in terms of the virtues of teachers’ assessment, but in terms of required positive impacts and assessment purposes leading to the identification of systems that support these impacts and purposes.

Paul Black suggested that thinking in terms of trade-off between competing properties of sub-systems (or models) was more appropriate than prioritising. He proposed describing sub-systems in terms of purposes and properties:

Purposes:	learning decisions accountability
Properties:	reliability validity comparability integrity backwash resources

Paul suggested that these might be used as criteria to describe some examples of teacher-based summative assessment sub-systems (models), such as the English course work GCSE operating in the early 1990s, the graded assessment schemes in the 1980s, the Queensland certificate, and the present system for comparison. This kind of analysis would make the case for using teacher assessment, an alternative to the approach advocated by Paul N. The issues raised were taken forward into the final session.

3.5 Further project action

Before beginning small group discussions Wynne introduced a framework for describing systems drawing on Paul N’s and Paul B’s contributions. The intention was not to describe all possible or feasible systems (as in the case of the model in WP1 Figure 3) but only those that could meet some view of ‘good’ summative assessment by teachers. It could be used within the argument ‘if you want to assess

these kinds of outcomes for these purposes, then these are the alternatives, and these are their advantages and disadvantages’. The bare outline (requiring a good deal of refinement) suggested was:

Profile		System 1	System 2	etc	
Use	internal external accountability				
Source of evidence	internal/on-going external bank external task				
Basis of judgment	progressive criteria task specific mark scheme (external)				
Training	internal (to the school) external				
Moderation	conducted internally/accredited status external agency				

This conflated two aspects in the WP1 model (criteria applied and form of judgement) and it was suggested that these may need to be disentangled again. However, details were considered within groups, where the task was to discuss whether this enabled the project to identify one or two ‘good’ systems which could then be discussed in terms of the properties identified by Paul B. It could also be used to describe existing, or past, systems in the same way.

3.3 Main points in reports from groups and subsequent discussion

- It was necessary to add to the model: the form of communication of the assessment (holistic or atomistic); the nature of aggregation (profile, cumulative, compensation); the source of the evidence and the timing of the judgments (modular or end-loaded).
- It was clear that high stakes accountability would distort the use of any assessment.
- Moderation should be seen as quality assurance in relation to all aspects
- It was considered preferable to use current example of practice in other countries (eg Queensland, Sweden, Scotland) rather than historical ones
- It was necessary to deal with discrete sub-systems (eg GCSE) within a whole system within an overall purpose; different sub-systems may be needed for specific domains.
- The project should have a mission statement declaring its principles in relation to purposes of assessment.

3.4 Further development

It was agreed that a draft document (WP3?) would be produced for discussion at the July Seminar, which would include an extended time for Core Group discussion. This would include: a mission statement, an identification of various systems and subsystems that met explicit criteria, and a discussion of the properties and pros and

cons of each. That would enable the project to take the next step of considering the extent to which we can identify 'how to do it'. Decisions about consultation/dissemination in the autumn would also need to be taken.