FOUR SURVEYS AND AN EPITAPH: AAP SCIENCE 1985-1997

RAE STARK, TOM BRYCE AND DONALD GRAY

SYNOPSIS

Four surveys of achievement in science have been completed to date under the auspices of the SOEID's Assessment of Achievement Programme in the last twelve years. These have taken place against a background of considerable change in educational policy, not least the introduction of the 5–14 Development Programme. This paper reflects upon these changes, the ways in which the AAP Science team has responded to them and what the findings have to say about science in Scottish schools over the years of the surveys. In addition it looks to the future and the need to monitor the implementation of the science guidelines and contribute to the essential work of increasing the effectiveness of learning and teaching in Scottish schools. This 'epitaph', we hope, marks only the end of an era and not the demise of what we have always believed to be an appropriate and defensible approach to monitoring standards in science.

INTRODUCTION

All four surveys of performance in science in Scottish schools undertaken within the Assessment of Achievement Programme have been conducted by teams of researchers at the Faculty of Education, University of Strathclyde and two of the authors, Rae Stark and Tom Bryce, have been involved in all four. This paper reflects upon the experiences of these years and, in particular, comments upon the increasingly political context in which national monitoring is undertaken.

The report of the fourth AAP Science survey, recently submitted to the Scottish Office Education and Industry Department (SOEID), was the first to attempt to measure performance in science against the expectations of the 5–14 Development Programme (Stark, Gray, Bryce and Ellis, 1997). It will be followed, as in previous years, by the publication of a summary report of the findings and a feedback document which will consider the implications for learning and teaching science 5–14. These publications will emerge shortly after the findings of the international surveys of science and mathematics (SOEID, 1996; SOEID, 1997). The relationship between national figures and international comparisons is an interesting one and deserves greater attention than can be given here.

FEATURES OF THE FOUR AAP SCIENCE SURVEYS

All four surveys have had the same basic aims:

- to provide information on what pupils at Primary 4, Primary 7 and Secondary 2 know and can do in aspects of science;
- to provide evidence on standards of performance over time; and
- to investigate ways in which the findings can be used to improve teaching and learning.

In order to ensure continuity and so determine trends over time reliably, many of the assessment materials and procedures were identical from survey to survey. Nevertheless, it was also necessary to introduce new elements to each survey in response to policy changes over the years.

A BRIEF HISTORY

The first survey, begun in 1985, was essentially a feasibility study which had two main aims. The first was to determine whether written and practical assessments could be carried out on a national basis (the logistics) while the second was concerned with developing appropriate sampling and statistical analysis procedures (the technicalities), drawing on the experience of earlier surveys in mathematics and English language. Some of the main features and findings of all four surveys appear in Figure 1.

In 1987, the context was relatively relaxed compared with the present, and the political stakes were low. Before the introduction of curricular guidelines and national testing, international surveys and performance indicators, teachers did not see monitoring as a threat and the AAP adopted a minimal intrusion approach with no teacher involvement in task development and marking, although, as with all AAP surveys, schools were asked to administer the written component. (Assessors were trained to administer the practical components.)

This first AAP Science team, led by Colin Holroyd, had a firm commitment to practical assessment: science was viewed as an active hands-on area of the curriculum and the assessment package used had to reflect that. This was costly, both in terms of materials and apparatus and in the training and payment of assessors to undertake the work in schools. The precise cost was difficult to estimate at the outset and the Scottish Office provided additional funds during the survey, ensuring its success.

The first survey was completed on time and successfully established reliable and valid procedures for science monitoring. Subsequently, the second and third surveys were commissioned from Jordanhill, the SOEID apparently recognising value for money and wishing to maintain continuity of experience. The fourth survey was put out to tender and, due to a lack of competing bids, it returned to Jordanhill, by then the Faculty of Education at the University of Strathclyde.

A number of changes took place during the twelve years of science monitoring, the most significant being the 5–14 Development Programme. Driven by concerns over standards of learning and teaching in schools, the publication of national guidelines for all areas of the curriculum, accompanied by arrangements for national testing, set out the government's expectations for science education across the primary stages and into the first two years of the secondary school.

Science was viewed as a component of Environmental Studies 5–14 (ES 5–14), the last of the five key subject areas to appear (SOED, 1993). National Testing was introduced in mathematics and English language (the two other subject areas regularly assessed within the AAP), but not in science. This is in stark contrast to the situation in England where science is a core subject area, alongside mathematics and English, and is included in national testing arrangements.

National monitoring in Scotland preceded national testing and differs significantly in both form and procedures in a number of key areas. Firstly, national monitoring employs a light sampling approach to pupil involvement, drawing nationally representative samples of pupils at each stage (approximately 2% of the population). Secondly, the assessment materials are designed to sample across the range of knowledge, skills and processes within a particular curricular area. It aims to provide a 'picture' of performance at each stage sampled. National testing, on the other hand, is concerned with providing information on performance and progress at the individual level and all pupils are likely to be tested at various stages in their school careers. National test materials tend to focus on a relatively limited range of concepts, skills and processes. The two systems, both concerned with national assessment, serve different but complementary functions, each being designed to meet a specific purpose.

Figure 1: Main features and findings, 1985-1997

AAP SCIENCE 1: 1985-1988

Feasibility issues: (i) the logistics;
(ii) the technicalities.

Context: pre-5–14, low stakes.

Response: schools — welcoming, friendly, interested.

Findings: P4 *compared favourably.*
P7 *performed well in unfamiliar situation.*
S2 *relatively poor performance.*

By gender: no significant pattern at P4; at P7, some evidence of girls > boys (skills); boys > girls knowledge and understanding (KU); at *S2*, boys > girls, (KU — physical sciences).

Across stages: P7 significantly better on all counts than P4; S2 significantly better than P7 on most tasks; P7 better than S2 on one or two tasks.
P4 and P7 more motivated and willing to try.

Schools: Significant differences in the goals of primary teachers and secondary science teachers.

AAP SCIENCE 2: 1989-1991

Based substantially on the work of the first survey, with a few minor changes.

Context: early 5–14, national testing, Working Paper 13.

Response: schools — cautious, but welcoming, friendly, interested.

Findings: P4 *satisfactory.*
P7 *satisfactory.*
S2 *performance was thought to be below what might reasonably be expected.*

By gender: girls and boys did equally well overall.

Across stages: differences between P7 and S2 less clear cut; more so between P4 and P7.

Across surveys 1987-1990: no discernible trend at any of the three stages assessed.

School policy: confidence of (primary) teachers.

and provision: seen as a major obstacle; some good examples of primary-secondary liaison in science.

AAP SCIENCE 3: 1992-1994

Context: environmental studies 5–14, published in 1993.

Response: schools concerned, but co-operating, generally. National testing, curriculum change.

Findings: P4 *largely satisfactory.*
P7 *largely satisfactory.*
S2 *performance at S2 was disappointing.*

By gender: some differences, which seemed to alter with age.

Across stages: some concern at the extent of overlap in performance between the stages.

Across surveys 1990-1993: no discernible trend at any of the three stages assessed.

Other aspects: collaboration with AAP English language team.
co-operation with Harlen, Holroyd & Byrne project, SCRE.

AAP SCIENCE 4: 1995-1997

Context: environmental studies 5–14 should be up and running!

Response: schools concerned: workload, initiatives, national testing, performance indicators, league tables.

Findings: P4 *satisfactory, although signs of weakness in specific areas.*
P7 *just satisfactory, although signs of weakness in specific areas.*
S2 *performance levels give cause for concern.*

By gender: clear patterns of gender-related differences.

Across stages: differences in P7/S2 performance does not reflect two years specialist science teaching.

Across surveys 1993-1996: improvement at P4; holding steady at P7; decline at S2.

Other aspects: pupils' attitudes to learning science.
collaboration with Gaelic project (Stirling University/Leirsinn).
modified re-run of Harlen *et al.* project.

Schools have been slow to implement ES 5–14, with around 20% of the primary and secondary schools in the 1996 survey describing them as 'fully operational'. Thus the 1996 findings can only be said to reveal a baseline of pupil attainment against the 5–14 targets but, unless there is a significant increase in the rate of implementation, particularly at the S1 and S2 stages, it is anticipated that performance levels are likely to increase very slowly indeed.

For the first survey, the framework was based on an analysis of existing policy documents and classroom materials and drew heavily on the work of the Assessment of Performance Unit's Science team in London. It was also informed by the experiences of team and Steering Committee members and the outcomes of research projects such as TAPS (Bryce, McCall, MacGregor, Robertson and Weston, 1991). While minor modifications were made over the years, significant changes were required for the 1996 survey in order to reflect the expectations of ES 5–14. This has tied the AAP much closer to the curriculum than before and has raised the (political) stakes for participants. (This was reflected in the reduced participation rate in the 1996 survey and the subsequent need to draw on additional, reserve schools.)

A SUMMARY OF FINDINGS AND TRENDS

The pattern of findings over the first three surveys was pretty consistent. Primary 4 and Primary 7 pupils regularly turned in performances which were judged to be satisfactory, indeed impressive, given the very varied science provision across Scotland. In 1996, when performance was measured against the expectations of the 5–14 Development Programme, P4 pupils attained Level B on the majority of the tasks which they tackled and P7 came close to attaining Level D targets. However, such broad brush statements mask substantial variation across the Strands and a poor grasp of some specific concepts, particularly those drawn from the *Energy and Forces* Attainment Outcome.

Secondary 2 pupils have been consistently disappointing in the standards achieved. In 1987, we reported on the *'relatively poor performance of pupils'* at this stage and in 1990, *'at S2, (performance) was thought to be below what might reasonably be expected'*. In 1993, the verdict was *'S2 performance ... did not seem as good as would have been expected, bearing in mind the two years of specialised teaching since P7'*, and in 1996 it was summed up as *'poor'*, with performance rarely better than Level D (approximately 15% of pupils regularly attained Level E targets).

While S2 pupils did, in each survey, outperform the P7 pupils, who in turn outperformed the P4 sample, the margin between the two older groups was not great, giving further cause for concern. Indeed performance at S2 appears to be in decline, with the performance levels of boys falling to a greater degree than girls. (Gender related trends in performance have become more evident with each survey.)

These findings are in line with the performance levels of Scottish pupils in science in relation to other countries as reflected in the findings from the Third International Mathematics and Science Survey (TIMSS: SOEID, 1996; SOEID, 1997).

THINKING POSITIVELY

The 5–14 guidelines are, fundamentally, a learning and teaching framework rather than an assessment framework. The AAP Science project has turned the guidelines into clear categories and sub-categories of knowledge and understanding, skills and processes, as well as providing exemplars of the kinds of tasks which are effective in their assessment. While the targets are deliberately broad in order that teachers can more readily meet the needs of their pupils and exploit their immediate environment, this lack of precision poses problems when attempting to define precisely what knowledge or skill the pupil should be able to demonstrate.

The 1996 AAP Science project put numbers against the 5–14 terms such as 'most' and 'almost all' (75% and 90% respectively) and determined cut-scores for 'secure understanding' and 'steady progress' in conceptual development (indices of achievement of 0.8 and 0.35 respectively). These figures were fixed upon in consultation with HMI and the Central Support Unit (CSU) at SCRE, amongst others, and while traditional test theory sets mastery at 80% of pupils achieving 80% of the available marks, it was decided that this would be somewhat stringent in the circumstances. These AAP Science figures now form the baseline measures against which any future monitoring exercise should be compared. Procedures for aggregation within Strands have also been established. (For details of how these figures and procedures were established and used, see Stark, Gray, Bryce and Ellis, 1997.)

In summary, such matters contribute to an understanding of what targets and levels may mean and how we respond to questions such as 'Have P7 pupils attained Level D in Science?' (In 1996, they did not — but nearly!) The significance of these matters is not lost upon members of HMI Science nor amongst members of the SCCC where there is considerable concern over standards in science in Scottish schools and a recognition of the role and contribution to the debate which the AAP provides. Unfortunately, this debate has not really been taken to the practitioners in schools who have expressed a need for more support in assessing and reporting performance in science (Harlen, 1995).

As members of the AAP Science team we have been, variously, involved in policy working groups and contributed to the dissemination of AAP findings from San Francisco to Sydney. While the science project has responded to political concerns and shifts in policy, it has not in the past been driven by them, although as the political stakes are raised, the desire for 'political correctness' in the Programme has grown. We believe that this has not compromised the quality of the work.

ON THE OTHER HAND

National monitoring has its limitations. While it provides a picture of what is happening in science education, it cannot be used to determine cause and effect relationships. It would be very interesting to be able to link performance to, for example, different teaching methods and materials, resource-based learning or class size effects but our instruments are not sufficiently probing. Nor can it explain why some of the findings are as they are — e.g., why do S2 pupils regularly turn in disappointing performances?

Sally Brown has been critical of the AAP in this regard, acknowledging its research limitations (Brown, 1991), while still recognising that the strategies used are 'fit for purpose' and meet the aims of national monitoring (Brown, 1994). So while we would argue that the AAP Science surveys have produced a good fix on national standards, they cannot satisfactorily shed light on the factors which contribute to the observed performance levels. The AAP can however pinpoint areas which would benefit from more focused investigation such as gender-related differences.

There have been other negative elements. The four main surveys have been so demanding that there has been insufficient time, energy and personnel to look in detail at specific aspects of the findings. Were the data to be reworked, more could be obtained on, for example, pupils' misunderstandings of concepts. The data is rich in this respect but further investigation demands a re-examination of pupil scripts rather than a secondary analysis of existing data.

In addition, while the funding allocated has been numerically similar from survey to survey, in real terms it has diminished with each, despite the increasing scope of the exercise.

SUMMARY

The four science surveys conducted by us have been conducted outwith the Scottish Office in a higher education institution with, arguably, significant research strengths and technical capabilities. We have received considerable support and advice from the SOEID and HMI over the years, as well as valuable contributions from numerous practitioners in schools and those serving on project committees, and for that we are grateful.

However, we are concerned for the future and would argue that the more significant contributions that have been made by the AAP Science surveys over twelve years of monitoring include:

- the conceptualisation of science education as a range of knowledge, skills and processes within an assessment framework which has relevance for teachers, advisers and policy makers as well as researchers and academics;
- the development of a wide range of testing formats, modes and strategies which have been tried and tested (and occasionally discarded) over the years;
- the development and mass production of practical forms of assessment, in large part due to the technical assistance provided (often free of charge) by the Science Division at Jordanhill;
- the collaboration with others, including the AAP English Language team and SCRE, as well as more transient relationships with Gaelic education and a number of international surveys;
- the contributions which members of the team have made to pre-service and in-service teacher education, at both undergraduate and postgraduate levels; and
- the publication of findings in journals and at conferences, often involving analysis of data beyond that required by the SOEID.

We would argue that recent changes in the organisation of the AAP are likely to weaken such contributions in the years to come.

The SOEID has reorganised the management of the AAP, appointing a national co-ordinator based in the Scottish Office who is responsible for the organisation and administration of all three subject areas. Amongst the co-ordinator's responsibilities is that of liaising with local authorities, each of whom has identified a member of staff to assist in co-ordinating AAP activity within the authority, with a positive impact on participation rates in the most recent mathematics and English language surveys.

In addition, within the lifetime of each survey, subject specialists are contracted, on a short-term, part-time basis, to provide advice on relevant aspects of the survey, including the development of assessment materials. Such individuals, typically 2 or 3 per subject area, are less likely to be regarded as constituting a 'team' with a committed institutional base and significantly independent of the SOEID, as has previously been the case (c.f. Bryce, 1996).

Our direct involvement with the AAP ended in July 1997 and with it, we would argue, an era of science monitoring which has been dynamic and positive in respect of its influence in Scottish education. We do fear that the reduced role of subject specialists and researchers, which appears inevitable given increased centralisation of the AAP, will result in a narrowing of the coverage within a survey, conveying a distorted picture of what science education should be. There is concern that science assessment will develop into a pencil and paper exercise, assessing what is cheap and quick to test (mainly knowledge and understanding) but we hope we are being unduly pessimistic.

Technically an epitaph belongs on a gravestone and we do not, as yet, wish to predict such a fate for the AAP Science programme. Nevertheless, having nurtured, developed and, occasionally, sheltered and protected it from occasional malevolence, we would wish to acknowledge the passing of an era. And so, with apologies to W. H. Auden, we have prepared an epitaph — just in case:

It was our North, our South, our East and West,
Our working week and our Sunday rest,
Our noon, our midnight, our talk, our song;
We thought it would last for ever; we were wrong.

NOTE

The views expressed here are those of the authors and do not necessarily reflect those of the SOEID who fund and manage the AAP.

REFERENCES

Brown, S. (1991), 'Research and Assessment: its influence on policy and practice', in *Assessment for the 90s. Seventh Forum on Educational Assessment in Scotland,* Edinburgh, S.C.R.E.

Brown, S. (1994), '5–14: Assessment and Testing', in G. Kirk and R. Glaister (Eds), *5–14: Scotland 's National Curriculum,* Edinburgh, Scottish Academic Press.

Bryce, T. G. K. (1996), 'Dependence Day for Data', Platform article, *Times Educational Supplement Scotland,* 27.9.96, 22.

Bryce, T. G. K., McCall, J., MacGregor, J., Robertson, I. J., and Weston, R. A. J. (1991), *How to Assess Open-ended Practical Investigations in Biology, Chemistry and Physics,* Heinemann Educational, Oxford.

Harlen, W. (1995), *Putting 5–14 in Place: An overview of the methods and findings of the evaluation 1991-95, Interchange,* 35, Edinburgh, SOED.

Harlen, W., Holroyd, C., Byrne, M. (1995), *Primary Teachers' Understanding of Concepts in Science and Technology,* Edinburgh, SCRE.

SOED (1993), *National Guidelines: Environmental Studies 5–14,* Edinburgh, SOED.

SOEID (1996), *Achievements of Secondary 1 and Secondary 2 Pupils in Mathematics and Science. Third International Mathematics and Science Study,* Edinburgh, TIMSS.

SOEID (1997), *Achievements of Primary 4 and Primary 5 Pupils in Mathematics and Science. Third International Mathematics and Science Study,* Edinburgh, TIMSS.

Stark R., Gray D., Bryce T. G. K. and Ellis S. (1997), *Assessment of Achievement Programme: Science. Report of the Fourth Round (1996),* Glasgow, University of Strathclyde.