

How do schools in Scotland measure their own progress?

4. Issues relating to the quality of data and analysis

In recent years, the Scottish education system has had an increasing focus on the measurement and management of performance by schools. As part of AERS, we are investigating the ways in which performance data are used by schools. Six schools in different areas of Scotland provide case studies, and demonstrate how they are using different types of data to inform practice and school improvement. In this Feedback we discuss some of the issues raised by schools about the quality of available data and measures of school performance.

- Assessment data are a key source of information about performance by schools and pupils. There are evident tensions between the different purposes for which assessment data are used.
- All of the primary case studies use baseline assessment to get to know their pupils' capabilities. None use the baseline data for subsequent tracking of pupils' progress or value-added analysis.
- All primary schools are required to carry out 5-14 National Assessments for accountability purposes, but many teachers consider them to be inappropriate and unreliable. Some schools prefer to use standardised tests.
- Perceptions that National Assessments do not provide appropriate measures of attainment are particularly problematic at the transfer from primary to secondary school.
- All secondary schools are required to evaluate performance using Standard Tables and Charts (STACS). These are summative data derived from National Qualifications (NQ) and aggregated at school or course level. The design of STACS does not include analysis/tracking of individual pupil progress – or value-added – between S1 and S4.
- The focus on National Priority measures creates perverse incentives for schools to target resources to particular pupils because they may improve the school's performance indicators.
- The proposal to extend the Scottish Survey of Achievement to all children would provide better data, but might have negative consequences.

Introduction

The first three AERS Feedbacks have focused on how schools use data for different purposes. For most purposes we found that key measures of performance are derived from assessment systems – especially 5-14 National Assessments and National Qualifications (NQ). Although schools and authorities do draw on other types of data – such as attendance and behavioural referrals, and aspects of wider achievement – the emphasis on easily-measurable attainment data overshadows other types of data.

There are evident tensions in the need for assessment data for different purposes. Teachers need to use assessment for formative purposes – to diagnose problems and identify next steps in learning – and in some schools we found enthusiasm for the approaches developed by the Assessment is for Learning programme. However, for tracking pupil progress a more objective type of assessment is needed in order to compare the relative progress of individual pupils over time – standardised tests are more appropriate for this

purpose but reduce an element of teachers' autonomy. In addition, accountability systems require summative data from external assessments to compare and evaluate how well classes and schools have performed – external assessments provide teachers with the least scope for professional judgements.

Interviews in the case study schools made it clear that national systems of assessment and qualifications are not adequate to support all of these purposes.

- National Assessments at the 5-14 stages are not helpful for diagnostic or formative purposes, and their binary measurement makes them inadequate for the purpose of tracking the progress of pupils from initial entry through the subsequent stages of schooling. At the same time their discretionary timing, subjective assessment, and vulnerability to “teaching to the test” make them of questionable validity for comparing performance between classes or schools.
- Some schools and local authorities use additional standardised tests – which increases the assessment burden on pupils and teachers.
- Current accountability systems are based on crude comparisons of performance between classes and schools. The absence of valid measures of prior attainment makes it difficult to develop more appropriate measures of “value-added” in terms of the relative progress of pupils between year stages.

To some extent these problems arise from political controversies in the 1990s – when teachers and parents successfully opposed the imposition of standardised testing similar to the national curriculum tests in England. The system of national testing introduced in Scotland relies to a greater extent on teachers' judgements but merely confirms whether or not a pupil has reached an expected level – this form of testing does not lend itself to analysis of pupils' progress.

Baseline data

A survey of local authorities in 2005 (Croxford & Cowie 2005) revealed that half used some form of baseline assessment at the start of primary 1. Baseline assessment has two purposes: (1) to provide teachers with information about the levels of readiness for learning, and to identify children who may have additional support needs; (2) to provide a baseline against which future progress can be measured.

We found that forms of baseline assessment were used in all three of the case study schools with primary classes – two carried out standardised assessments of reading and mathematics, and one used the more qualitative nursery/primary transfer record. In each case teachers made use of the baseline assessment in order to get to know the

pupils' capabilities, and to identify those who might need to be assessed for additional support needs. However, none of the case-study schools used the baseline data for subsequent tracking of pupil progress – possibly because of the lack of calibration of baseline tests with subsequent assessments.

5-14 National Assessments

Currently, 5-14 National Assessments are the main source of data for evaluating performance in primary schools. Until recently the National Assessment results were used by the Scottish Government to set targets for improvement in performance of primary schools – and some local authorities still use them for this purpose. However, the data proved so unreliable that they are no longer collected on a national level.

In one of the case study primary schools efforts are made to use 5-14 data to track pupil progress. Pupil-level histories of National Assessments, recorded in the management information system, show the stage and date at which each pupil has achieved each level in reading, writing and mathematics. Potential problems are identified if there is a longish gap between achievement of each level. Every three months the class teacher and deputy headteacher meet to discuss progress of individual pupils – based on historical data – to set targets for their next levels and to provide additional support for the child or the class where necessary.

The headteacher believes that this tracking system enables her/him to keep an eye on overall progress and intervene if necessary.

I can track it very clearly, and I know the areas that children – the class as a whole or individual children – have particular strengths in, particular successes or indeed specific weaknesses that perhaps need picking up or addressing... It's not a case in this school of teaching towards the test. What we do try and do is to accelerate things slightly, but that's not a bad thing. What we are trying to do is to get the children to reach their full potential, and you do that in a number of ways.

In other case study schools National Assessments are widely condemned as unreliable evidence (although the schools continue to carry out the assessments because of pressures to comply with the accountability system).

One primary school is so dissatisfied with the system of 5-14 National Assessments – especially their inadequacy for identifying additional support needs – that they use commercially-produced standardised tests with their pupils on a regular basis.

(The standardised tests results) come alive when you have got a lot of experience in applying them to real children – to real learners in classrooms. I can't say I have that with national assessments. I don't know what it is. They are the strangest set of assessments I have ever seen. They are uniquely strange. I know that they have got a history of criterion referenced assessments but... certainly there are children who come here who have level A in this or level A in that and... as soon as you have done a standard assessment you know the child is not performing at that level. I don't know how they get these results"

The school finds these standardised tests very effective in identifying additional support needs.

"... It is the testing regime. ... it is because children are regularly measured. A child can wait a long time for a [national] assessment."

However, at present the school does not have an appropriate pupil tracking system, and does not make full use of the standardised test data to track pupil progress over subsequent stages.

Problems at primary/secondary transfer

Perceptions that National Assessments do not provide appropriate measures of attainment are particularly problematic at the transfer from primary to secondary school. One school has attempted to improve the quality of assessment through a moderation exercise.

Primary teachers said "we gave secondary all this good data about children and they don't use it – they just say 'we'll start afresh' ". Secondary teachers are saying, "well we get this broad band – what does that mean?" So we start some moderation work between secondary and primary. "Here are your national assessments, mark them, particularly for Language. Are there discrepancies between how Secondary perceive a level D in Primary?" – and there were. Primary believe Secondary are marking too hard, and I tend to agree with them. ... And we've done that with our cluster, and we intend to do more of that. That's been well-received by teachers. I think that gives staff faith that when schools in any of these connections say it's a level D – it's a level D.

Disenchantment with National Assessment data has led two of the secondary case study schools to introduce cognitive abilities tests (CAT) with their S1 pupils to provide more robust baseline information. (In one case

this is at the behest of the local authority which now uses CAT at a number of stages). CAT includes three principal areas of reasoning – verbal, non-verbal and numerical – as well as an element of spatial ability, and gives predictions of Standard Grade performance.

One school uses CAT data at S1 and S3 as part of a pupil-tracking system, and for very detailed personalised target-setting (See AERS Feedback 1). However, staff need to recognise that CAT scores and predictions may not be entirely reliable.

What the CAT testing data has been great for is giving them a baseline and giving them a sense of, or helping them more quickly to establish a sense of where these pupils should be. And I see the data as largely accurate – there are the odd exceptions in it. But I would say broadly, the CAT testing data seems accurate – it's quite close to being accurate, and quite often when a pupil's not achieving it's for an obvious reason – very poor behaviour, lack of motivation. So we will use that data in conjunction with our teacher's data to set the classes and to establish mixes of pupils. We also use that data to assist with target setting and so forth.

Limitations of data in the secondary stages

Currently, STACS are the main source of data for evaluating performance in secondary schools, as described in AERS Feedbacks 2 and 3. STACS have two main weaknesses: (1) they are summative data after pupils have completed national courses; (2) they focus on aggregate data at the level of departments and school and do not provide pupil-level data. Some schools and local authorities buy-in additional pupil-level information from consultants but, in the main, STACS do not support schools in tracking the progress of individual pupils.

The narrow focus of STACS ignores the attempts by schools to provide wider achievement recognition for pupils with less chance of academic attainment. These include Skills for Work courses, ASDAN, Duke of Edinburgh's Award, Sports Leadership Awards, and John Muir Award (an accredited environmental award).

Although the validity of NQ data is appreciated by schools and local authorities, the methodology used by STACS does not provide the most appropriate measures for evaluating performance. Value-added measures that take account of prior attainment and estimate progress over a number of year stages would provide a fairer way of evaluating school and subject performance than comparisons based on a single time-point. STACS provide measures of value-added (ie Progression Values) between S4 and S6

using S4 Standard Grade as a baseline for estimation of subsequent progress. However, prior to S4 the system does not have a consistent measure of prior attainment so STACS use concurrent attainment to create a measure of “relative value”. This measure is flawed because it cannot measure relative progress through prior stages. Its usefulness is further reduced by the increasing use by schools of NQ other than Standard Grade.

Currently, inspectors and other external observers judge how well the school is doing in comparison with others by National Priority measures of attainment, such as the percentage of each cohort achieving five or more awards at Levels 3-6 of the Scottish Credit and Qualifications Framework (SCQF).

The whole school measure ... that's what's going to appear in the papers – that's the benchmark figure against which we'll be judged by the authority and by the Inspectorate.

The focus on National Priority measures creates perverse incentives for schools to target resources to particular pupils because they may improve the school's performance indicators – for example pupils who are likely to achieve just four awards at an SCQF level, but who may – with extra support – be able to achieve five awards. This practice appears inequitable, because it may help some individuals to achieve more highly, while other lower-attaining pupils do not get the extra support (Cowie et al 2007).

Recent and future developments

The case studies for this project were carried out in spring 2007, and subsequently there have been a number of policy developments, such as the provision of greater autonomy to local authorities, which may impact on the use of data. The recent OECD Review of Scottish education commented on the need for more reliable evidence about educational standards across Scotland. The review suggests that the Scottish Survey of Achievement (SSA), should be extended to assess all children throughout Scotland.

However, this recommendation raises concern about increasing the burden of testing in schools. Consultation with parents and teachers in the National Debate showed a widespread desire for less testing, and as a consequence the current focus of initiatives such as Assessment is for Learning is on formative assessment. The SSA is currently based on a representative sample of pupils and schools, but already places a burden of summative assessment on pupils and schools. To extend the burden to all pupils could have negative effects on learning and teaching.

A Curriculum for Excellence, and the related consultation by the Scottish Government regarding the next generation of NQ, raise many issues for schools to address in the future, including increased flexibility

in the timing and duration of NQ courses, and the proposed tests of Literacy and Numeracy at the end of compulsory schooling (Scottish Government 2008).

There also appears to be a significant shift in HMle's approach to inspections. In the past, schools and authorities knew that inspectors' judgement of a school overall, and its leadership in particular, would depend almost exclusively on how well the school was attaining. HMle now suggest the quality of self-evaluation will have a more important role in future inspection, and that the school's own view of its performance overall will determine the grade they give. Furthermore, the judgement is to be based on not just attainment but on a number of Quality Indicators, thus giving scope for consideration of wider aspects of achievement. Schools are urged to use their initiative to be 'data rich' – with the implication that data regarding a whole range of aspects of pupils' learning and wider school experiences will be expected.

References

Cowie, M., Taylor, D., and Croxford, L. (2007) 'Tough, Intelligent Accountability' in Scottish secondary schools and the role of Standard Tables and Charts (STACS): a critical appraisal, *Scottish Educational Review*, 39 (1) 29-50.

Croxford, L. and Cowie, M (2005) Accounting for School performance in Scottish Local Authorities, *CES Briefing No 37*

OECD (2007) *Reviews of National Policies for Education: Quality and Equity of Schooling in Scotland*

Scottish Government (2008) *A consultation on the next generation of NQ in Scotland*, Edinburgh: SG

Findings of this research

AERS Feedbacks to case-study schools comprise:

1. Tracking systems
2. Evaluating data to identify issues
3. Formal accountability processes
4. Issues relating to the quality of data and analysis
5. Equity issues arising from the use of performance data

CES Briefings 37 and 43 report other findings of this AERS study and can be downloaded from:
www.ces.ed.ac.uk/publications/briefings.htm

For further information please contact:

Dr Linda Croxford

Phone: 0131 651 6283

Email: L.Croxford@ed.ac.uk

Dr Michael Cowie

Phone: 0131 651 4190

Email: Michael.Cowie@ed.ac.uk

University of Edinburgh, Moray House School of Education, St John's Land, Holyrood Road, Edinburgh EH8 8AQ