

COMMON
GROUND on
ASSESSMENT and
ACCOUNTABILITY
in
PRIMARY
SCHOOLS



Common ground on assessment and accountability

Arguments about whether or not schools should be accountable for the quality of education they provide should be consigned to history. The question is not whether schools should be accountable, but how they should be held to account, and to whom.

Since the opening of the modern era of school management in England, with the passing of the Education Reform Act in 1988, the accountability system has mushroomed. Results of assessments – including national curriculum tests, GCSEs and A-levels – are central to that system. However, the data generated by these assessments are now used in ways which extend far beyond what might be seen as their most obvious function: providing a check on a pupil's understanding and progress in a subject. A detailed look at the purposes of the assessment-based accountability system, its uses and its problems, is long overdue.

The associations and organisations which have signed this document have both a common commitment to public accountability and a common set of concerns about the way it currently operates.

Their concerns can be summarised under five headings:

England's test and exam-based accountability system is excessive, confused and overlapping, to an extent which would not be tolerated in the private sector. This also provides poor value for money.¹

As the House of Commons Children, Schools and Families Committee put it in January 2010, "The complexity of the school accountability and improvement system in England is creating a barrier to genuine school improvement."²

There is a widely-acknowledged specific problem in relation to the use of assessment data for accountability. In trying to use the same data sets to serve many purposes all at once, it satisfies none of them very well.

There have been many attempts to categorise the number of uses to which assessment data are put. The most commonly-cited analysis suggests that there are 22. These extend from providing judgements on a young person's suitability for a particular course of further study or career to evaluating the quality of individual schools, of local authorities and of the nation's educational performance³.

These purposes are often in conflict.

¹ Estimates have put the cost of England's primary assessment system at around £20 million. This excludes the costs of Ofsted. Tymms P. and Merrell C, Interim Report 4/1 "Standards and Quality in English Primary Schools Over Time." University of Cambridge/Esmee Fairbairn, (2007).

² House of Commons Children, Schools and Families Committee: "School Accountability", First Report of Session 2009-10.

³ Qualifications and Curriculum Authority submission to the House of Commons Children, Schools and Families Committee inquiry into "Testing and Assessment", 2007/08 session.

In evidence to the Committee, Dr Ken Boston, then chief executive of the Qualifications and Curriculum Authority, said: "There are 22 purposes currently being served by current assessments, and 14 of those are in some way being served by Key Stage test assessments...when you put all of these functions on one test, there is the risk that you do not perform any of those functions as perfectly as you might. What we need to do is not to batten on a whole lot of functions to a test, but restrict it to three or four prime functions that we believe are capable of delivering well."

The select committee, commenting specifically in relation to national curriculum tests, concluded: “National tests do not serve all of the purposes for which they are, in fact, used. The fact that the results of these tests are used for so many purposes, with high-stakes attached to the outcomes, creates tensions in the system leading to undesirable consequences, including distortion of the educational experiences of many children ... We consider that the current national testing system is being applied to serve too many purposes.”⁴

A better system would seek to define what purpose each assessment is supposed to fulfil, and most importantly, what it is not, and then seek to construct assessments which were fit for the required purposes.

Accountability is hugely over-centralised. Schools rightly need to be held to account by parents, pupils, local communities and political decision-makers. However, the current accountability structure places far too much onus on schools being accountable to central government, at the expense of other stakeholders to whom they have a responsibility. Better arrangements would centre on the different kinds of information that schools should provide for the different needs of pupils and their parents, local communities and national policy-makers.

The current system, though finding a greater place for self-evaluation than it did in the past, is still predicated on a lack of trust in the teaching profession. Schools need to be allowed to move further towards self-evaluation, but subject to external scrutiny. Improving and supporting teachers’ capacity to make well-informed assessment judgements about the progress of their pupils – and to check on the assessment judgements of colleagues working in or beyond their schools – will be crucial. Schools work better when teachers and school leaders feel they can take ownership of the process of improvement, rather than having it imposed upon them.

Results generated by the accountability system are subject to unnecessary and damaging over-interpretation. A school’s overall performance can never be summed up in a single number, or even in a handful of central government statistics, but this is the implication of how performance tables and Ofsted inspections often work. School staff, especially school leaders, believe a single set of bad test results could damage or end their careers. This is hugely damaging to morale and energy, especially for those who choose to work in the most challenging environments.

Most importantly, the assessment-based accountability apparatus too often serves to undermine pupils’ learning. Evidence of this is copious. Independent research and reports by organisations including the House of Commons Schools Select Committee, the Qualifications and Curriculum Authority, Her Majesty’s Chief Inspector of Schools and leading academic bodies shows that the pressures on schools to raise results have had damaging side-effects. These have included pupils being taught a limited and unbalanced curriculum, particularly but not exclusively in year 6, as teachers feel constrained to tailor their teaching towards test preparation; the effects on children of the increased emphasis on assessment performance, including excessive stress felt by some pupils; and the concentration by schools on particular groups of pupils whose performance is on the borderline of success or failure under Government indicators.

⁴ House of Commons Children, Schools and Families Committee: “Testing and Assessment”, Third Report of Session 2007-8.

Improving the system

Reform of England's assessment-based accountability mechanism, then, is urgently needed. In the following section, this has been broken down into five categories, reflecting the detail of how accountability now works, and how it should work.

Testing and teacher assessment

Externally-moderated teacher assessment should replace national curriculum "SATs" tests as the dominant means of holding primary schools to account. Teachers should be seen as experts in assessment, as it is a core professional role. In order for teachers to have ownership of assessment, they must have control of it and be free from bureaucratic demands, external to the school, which distract their time and energies away from teaching.

Teacher assessment, when it is subject to proper external checks and is carried out by well-trained professionals, has clear advantages over national curriculum testing.

First, it is a more valid form of assessment, because a much wider range of pupils' learning, over a much longer time period, can be evaluated by the teacher than is possible through a few short, one-off, tests⁵. It thus more easily integrates assessment with the curriculum and pedagogy. The best teacher assessment should be built naturally into the normal activities of lessons.

Second, it can be at least as reliable as SATs in providing accurate judgements about the levels pupils have achieved. There is a myth that says that because national curriculum tests are marked by teachers with no connection to a school, the marking will be "objective" and free of error. However, there is copious research, including from test regulators, that SATs are subject to marker error⁶. Secondary schools also lack confidence in the results of SATs taken in year six, as demonstrated by the fact that a high proportion give pupils a fresh set of tests in the first term of year seven⁷.

SATs also have little value in giving pupils, and their parents, the information they need to support further learning progress. Because Key Stage 2 SATs take place towards the end of a pupil's time at primary school, with results released in the last weeks of their final term, the marks generated are of no use to teachers in planning each pupil's further learning needs.

By contrast, in-class teacher assessment can play a formative role in promoting further learning. This formative process is also known as Assessment for Learning. With appropriate feedback, pupils become more active and committed learners, and

⁵ This was acknowledged, for Key Stage 2 science assessment, by the Department for Children Schools and Families. *Teacher assessment*, it said, "takes greater account of pupils' practical grasp of the subject and is based on their attainment throughout the academic year across the full programme of study." Source: <http://www.dcsf.gov.uk/lrsgateway/>: "DCSF: Changes in the reporting of National Curriculum Assessments at Key Stage 2 in England 2010: Introduction of Science Sampling" Accessed 11th August, 2010.

⁶ See William D. (2001) "Level Best? Levels of Attainment in National Curriculum Assessment", in which it was estimated that at least 30 per cent of pupils could be misclassified in national tests. Newton, P (2008): "Presentation to the Cambridge Assessment Forum for New Developments in Educational Assessment", Downing College, Cambridge, 10th December, suggested a figure of 16 per cent. In 2009, the Qualifications and Curriculum Authority analysed how often markers agreed with each other on which levels to award pupils in the now-defunct key stage 3 tests for English, maths and science. The frequency of agreement was as low as 56 per cent for English writing. Source: Qualifications and Curriculum Authority (2009): "Research into marking quality: studies to inform future work on national curriculum assessment". London: QCA.

⁷ De Waal, A. (2008): "Fast Track to Slow Progress", Civitas.

their progress improves. Improving in-class feedback – the information passing between teachers and pupils on what the learner needs to do to improve – has been shown by research to be the single most effective way of improving educational performance⁸.

Unfortunately, there is also much evidence that the current system of high-stakes summative testing can hamper effective formative assessment, thereby undermining teacher professionalism⁹.

Teacher assessment judgements that are used to report formally on the progress of pupils, and of their schools, need to be subject to moderation by well-trained professionals. The previous government's attempt to standardise teacher assessment judgements, known as Assessing Pupils' Progress, has proved successful in some classrooms, but it contains the potential to be overly-bureaucratic unless its use is under the control of teachers. It has worked best, improving the quality of assessment, contributing significantly to teachers' professional development and helping secure sustainable high-quality learning, where professionals have been allowed to take ownership of the process within mutually supportive professional learning communities, rather than feeling it has been a centrally-imposed requirement. Too often, though, teachers have felt coerced into APP, reducing its effectiveness. It may be that there is no national "one size fits all" model for supporting teacher assessment, but that schools should be supported to work in partnerships with local authorities and with approved trainers working to nationally-accredited models.

Like all human judgements, teacher assessment will not be free of error. For example, research has shown that teacher assessment can produce results which are biased along ethnic lines.¹⁰ They must be given access to continuing professional development in teacher assessment, on an ongoing basis to address such issues. Teachers, though able to assess the national curriculum levels achieved by their pupils, are not always consistent in their own assessments or in comparison to other teachers. Moderation by other teachers, both within the school and in other schools, is therefore essential. There is a need to develop a cohort of teachers who are experts in assessment. This must be teacher-led and locally organised, but accompanied by a resource which supports national standardisation: a national bank of assessment materials from which teachers can choose to draw to check their assessments. Building teachers' expertise is vital. Teachers' growing mastery of assessment could be recognised formally through the development of a chartership or assessment champions.

Improvements in teacher assessment have the potential, then, to bring about major gains for pupils. These must not, however, be undermined by the imposition of additional burdens on schools and their staff.

⁸ See Hattie, J. (2009): "Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement" Routledge, or a report of Hattie's research at <http://www.tes.co.uk/article.aspx?storycode=6005411>

See also: Black, P., Harrison, C., Lee, C., Marshall, B and Wiliam, D (2002): "Working Inside the Black Box: Assessment for Learning in the Classroom", King's College, London and Wiliam, D. (1998): "Assessment and Classroom Learning", *Assessment in Education*, 5 (1), pp. 7-71.

⁹ See Black, P., Harrison, C. Lee, C., Marshall, B. and Wiliam, D (2004): "The Nature and Value of Formative Assessment for Learning", King's College, London;

Smith, C., Dakers, J., Dow, W., Head, G., Sutherland, M., and Irwin, R. (2005): "A Systematic Review of What Pupils, Aged 11-16, Believe Impacts on Their Motivation to Learn in the Classroom", London: EPPI-Centre, Institute of Education, University of London.

¹⁰ Burgess, S. and Greaves, E. (2009): "Test Scores, Subjective Assessment and Stereotyping of Ethnic Minorities". Centre for Market and Public Organisation, University of Bristol.

There is scope for a reduction in unwanted workload through the replacement of high-stakes testing as the main accountability mechanism for primary schools. Teachers would benefit from a reduction in the time spent on test preparation. In all schools, they would also gain if freed from the centralised control of the current accountability system: excessive workload comes from a feeling that aspects of their work which should be the locus of their professional judgement – curriculum, pedagogy and assessment – are being imposed from above.

National sampling

A system whereby a representative sample of pupils are assessed every year should replace SATs tests as the main measure of national education standards and the most important national accountability measure. Pupils' outcomes in such tests would not be linked to the schools they attended, meaning they would be "low-stakes" for pupils, teachers and schools.

Such assessments would have several advantages over SATs as measures of national standards. They are a good example of how developing an assessment model with one express purpose – in this case to measure national educational performance – is better than expecting one set of assessments to fulfil multiple purposes.

First, because they would be "low-stakes", test questions could be retained from year to year, so that direct comparisons could be made between pupils' performance on questions in successive years. This would enable a much more authoritative picture to be established on national educational performance than is currently possible.

Second, the much smaller numbers of children being tested mean that it would be practicable to measure children's progress and understanding on a much larger part of the curriculum than is possible now, and in a more sophisticated way¹¹.

Third, there would be less pressure on schools to "game" the testing system by focusing on particular groups of pupils on the borderline of achieving Government expectations, or to concentrate on particular types of question predicted to come up in the tests. Again, the sample tests, if well-designed, would then be expected to provide a better overall view of underlying trends in national education performance.

Fourth, the low-stakes nature of the tests would mean schools would not engage in excessive test preparation, so children's education would not be disrupted.

Such sample tests are well-established in other countries, and in international testing studies such as PISA, TIMSS and PIRLS.

Inspection

Inspection needs to be supportive of the school, becoming more of a shared professional experience between the school's leadership, its staff and the inspection team, rather than feeling like an external invasion.

¹¹ *The Assessment of Performance Unit, which measured national standards in the 1970s and 1980s, tested representative samples of pupils on scientific experimental work and spoken English and oral comprehension – which would be extremely difficult to measure through tests taken by a full national year group of pupils – alongside tests in mathematics, English, science, foreign languages and design and technology. See Alexander, R. (ed) (2009) "Children, their World, their Education: Final report and recommendations of the Cambridge Primary Review", Routledge.*

Self-evaluation, subject to external verification, should be the central principle. Self-evaluation leaves the improvement of the institution in the hands of the people best placed to drive that improvement: those working in schools, students and parents. The Government should avoid expecting one-size-fits-all approaches to self-evaluation from schools.

Inspection should be proportionate to the need for it: schools which are clearly of high quality should have less frequent inspections than those where there are concerns, with the very best subject to very light-touch quality assurance. However, inspections currently can be so short that there can be an undue reliance on data. This should be avoided. In addition, inspectors should avoid relying on single national data indicators, such as Contextual Value Added statistics. Schools use other statistical analysis systems, which can provide equally valid measures. This demonstrates the danger of narrowing accountability measures unrealistically. A single measure should never be seen as the sole “true” verdict on performance.

Surveys of pupils and parents are now a part of many schools’ self-evaluation processes. They should not, however, become subject to central control by the Government or Ofsted, but should remain under the ownership of schools.

In addition, schools should never be judged on the basis of a single year’s examination results. Results over at least three years give a much better guide to trends.

Performance tables

Parents do need to be provided with information on the progress and achievements of pupils in their child’s school, or in schools to which they are interested in sending their children.

However, the current league tables, based on too narrow a definition of pupil performance, which is particularly problematic at Key Stages 2 and 4, provide imprecise and misleading measures¹², while serving to distort the curriculum and devalue the creative and broader personal skills that are essential for successful and fulfilled lives.

Too often, the tables demean the work of schools and mistakenly categorise their qualities.

The central statistical measures used in compiling the tables – the proportion of pupils achieving the Government’s expected level in English and maths at Key Stage 2, and the proportion achieving five or more GCSE A*-C grades, including English and maths, at Key Stage 4 – act as perverse incentives for schools to concentrate on pupils at the borderline of achieving these indicators. They provide no encouragement for schools to tailor provision for children working above or below such expectations.

¹² See Goldstein, H. and Leckie, G. (2008): “School league tables: what can they really tell us?” Volume 5, issue 2 of the Royal Statistical Society’s magazine, “Significance”. They say: “The inherent imprecision of all estimates reduces their usefulness for accountability purposes. We have said nothing about the side-effects and perverse incentives generated by the use of league tables. These are undoubtedly serious.”

In “The Tiger that Isn’t: Seeing Through a World of Numbers” (2007), authors Michael Blastland and Andrew Dilnot conclude: “Ministers often said that league tables should not be the only source of information about a school, but it is not clear in what sense they contributed anything to a fair comparison of school performance or teaching quality.”

The Government's move to reform league tables to enable schools to demonstrate the progress of pupils of all abilities is therefore welcome. It should also review the Contextual Value Added system, the last Government's attempt to measure schools' contributions to their pupils' achievements, taking into account their backgrounds.

Indeed, it should go further. Especially as members of the new Government have talked about working in partnership with professionals¹³ and local communities, there is no longer any need for performance table information to be published.

Instead, schools should be required to make appropriate information on their performance available to parents, if necessary in a standard form with an agreed range of indicators. In primary schools, this would include information on pupils' progress as generated by teacher assessment judgements.

Accountability for school partnerships

Partnership arrangements between schools are becoming more common, with institutions now working together in federations, "chains" of schools being overseen by sponsors and new academies now expected to work with partner schools.

The Government should avoid imposing another form of accountability on these partnerships on top of existing arrangements. If the trend towards partnership continues, accountability will need to be adapted to provide a greater emphasis on the work of the group.

As far as possible, and subject to piloting, schools in partnership should be inspected together.

Conclusion

Assessment-based accountability has a great influence on what goes on in schools. Getting it right will accordingly bring great benefits to pupils, parents, policy-makers, the school workforce and the nation as a whole.

The best form of accountability will provide the information needed by all who have a stake in the quality of school education while demonstrating greater trust in the teaching profession, restoring sensible, externally scrutinised, professional autonomy to all those who work in schools. This has the potential to inject great energy and enthusiasm into institutional improvement.

We believe the changes set out in this document will produce strong public accountability but with far fewer negative side-effects for pupils, teachers and schools.

¹³ "We...believe that those most in need will never be helped to achieve all that they can unless we harness the full power of civil society, the initiative of creative individuals, the imagination of social entrepreneurs, and the idealism of millions of public sector workers. That means reducing bureaucracy, getting rid of misguided political intervention, respecting professional autonomy and working in genuine partnership with local communities." Gove, M., Queen's Speech Debate, June 2nd, 2010.