Summer 2010

Make Assessment Measure Up

Both the National Union of Teachers and ATL believe that the National Curriculum assessment arrangements are in urgent need of reform.

We believe that assessment involving all pupils should focus on enhancing their learning, not on evaluating schools. Other forms of evaluation should focus on institutional effectiveness. Sample tests should be used to help evaluate the education service as a whole.

This paper argues that the replacement of current national testing arrangements at Key Stage 2 by moderated teacher assessment, together with sample testing, would benefit pupils, parents, the Government, and teachers. It also outlines the necessary developments needed to put such changes into place.

NATIONAL UNION OF TEACHERS NUT

ATL the education union

Introduction

The two purposes of pupil assessment should be to support learning (formative) and to report achievement resulting from learning (summative). Formative assessment serves the purpose of promoting pupils' further learning. Summative assessment judges pupils' performance at a point in time. In order to achieve these aims, both forms of assessment need to be integral to the curriculum and to teaching and learning.

The terms 'testing' and 'assessment' are sometimes used interchangeably. This is an error. Testing is one method of assessing; like other methods it has strengths and weaknesses, and policymakers should evaluate tests against the other kinds of assessment discussed below.

Current Arrangements and their Weaknesses

Current assessment problems are not new, but are derived from the derailing of the work of the Task Group for Assessment and Testing (TGAT), set up by the Government in 1988 to establish the framework for the National Curriculum, including its assessment levels and programmes of study (PoS)[1]. Under pressure from Ministers the PoS were calibrated into subject content specific to each Key Stage. The TGAT's conception of an integrated curriculum which could be referenced for diagnostic purposes by teachers, accompanied by summative assessment solely at 16+, was abandoned, paving the way for the end of Key Stage tests with no comparability between assessments at each stage.

Assessment for Accountability

Currently, excessive weight is placed on the outcomes of pupil assessments. Any assessment tool, including tests, has to be designed for a specific purpose; for example, according to whether the intended purpose is either formative or summative. Yet, according to the former Qualifications and Curriculum Authority (which became the Qualifications and Curriculum Development Agency in 2009 but looks likely to be abolished under new proposals) test results are used for 22 different purposes[2]. Many of these, particularly those connected with accountability, are inappropriate. They are used to measure national pupil performance and its changes over time. Test results are also used in the so-called league tables, by Ofsted and by local authorities to evaluate schools. They have a high local and national profile and are subject to intense and often misinformed political scrutiny. The result is that school staff, and particularly school leaders, believe that a single set of test results might well damage or end their careers. This is despite the fact that, even within the narrow parameters of testing, evidence shows that five-year rolling results give a better picture of performance in the areas covered by assessments than one year's results. In such circumstances, risk averse behaviour such as teaching to the test and aversion to innovation is highly rational.

In short, there is now a large and excessive network of accountability mechanisms affecting schools and teachers. They include multiple methods of staff monitoring, performance management, local authority monitoring, School Improvement Partners and Ofsted, as well as the legitimate expectations of parents. School league tables have failed as a proxy for evaluating school quality. School accountability needs rationalising as a matter of urgency.

National Curriculum Tests and the Restricted Curriculum

Independent research, as well as successive Annual Reports from HMCI, have shown that many primary pupils, particularly but not only in Year 6, are taught a limited and unbalanced curriculum because teachers have felt constrained to tailor their teaching to SATs' test items[3]. This tends to produce 'surface' or 'shallow' learning, in which facts can be recalled but without deep comprehension. Recall deteriorates after the test, which is partly why measured pupil performance dips between Year 6 and Year 7[4].

Effects on Pupils

There is substantial evidence that the current focus on high-stakes tests has negative effects on pupils and their learning. Many believe test results have a higher value than what they have learnt. Testing has been found to be stressful[5] and demotivating[6], particularly to lower achieving pupils on whom there is now such a policy focus. While there are those who argue that the stress of testing is 'a good lesson for life', the reality is that excessive stress inhibits learning. Indeed, international evidence shows that the most educationally successful countries postpone national testing[7].

Summative Assessment by Tests

All of the above problems would be found in any assessment system where the stakes are as high as those which apply to the current testing arrangements. However, there are additional problems with National Curriculum Tests.

Both the Government's assessment agency, and its contractors, work hard to achieve validity and reliability in tests and exams. However, it remains true that both tests and public exam grades are not exact measures, but have known margins for error. No-one likes to shout about this for fear of undermining public confidence, but when an individual student is mis-graded at A level it can be a personal disaster. And, when the education system relies on Key Stage test results with an unreliability of 30 per cent[8], that is a policy disaster. Although errors may have a neutral effect (with over-grading cancelling out under-grading) in a national cohort of students, the effect is not necessarily neutral in a small sample, for example, a single Year 6 group.

Policy disasters occur when too much weight is put on test results, particularly where there is an implied assumption of 100 per cent accuracy. Ofsted inspection grades are strongly based on test results, with 'limiting judgements on attainment' discriminating against schools with socially disadvantaged intakes. Indeed, the RAISEonline database, which purports to compare schools with similar intakes, has methodological flaws. Retesting of pupils on entry to secondary school is widespread because of secondary schools' lack of confidence in the validity of SATs results[9]. Parents continue not to behave according to school market theory by placing little reliance on SATs results when choosing a primary school[10]. In general, parents are more concerned about their children, their children's happiness, security, and whether their progress is as good as it should be, and less concerned about how the school performs as a whole in tests. Government's support for parents should reflect this enduring emphasis.

Internationally, the most successful school systems test pupils least and latest[11]. A review of tests by the Daugherty Assessment Review Group[12] led to their abolition by the Welsh Assembly Government. The Peacock Review[13] in Scotland reached similar conclusions. England is exceptional. Policymakers in the Westminster Government need to consider carefully why only England is out of step.

There is, however, a strong argument for a national test to measure trends in national performance. As argued above, to be valid it has to be low stakes at school level. This would be achieved by the replacement of a test of the whole cohort by small sample tests. Using sample tests would also save most of the approximately £20 million spent on Key Stage 2 tests[14].

Undermining Teacher Professionalism

An underestimated drawback to the national test system is its effect on the stock of professional knowledge and skill within the teaching force. The most effective teaching requires the continuous interplay of formative assessment and lesson planning[15]. The skill of a teacher is to know what a pupil has securely learnt in order to plan the next steps in learning. However, there is evidence that the inappropriate emphasis on summative assessment because of end of Key Stage tests critically undermines formative assessment practices[16]. The grade focus of the former undercuts the improvement and task-oriented

focus of the latter. While teachers in England have become highly skilled at assessing the National Curriculum level of pupils, this is of little use for planning personalised learning.

Many teachers have a low level of confidence in their ability to make assessments independently. This is brought into sharp relief by the excessive prominence given to test results alone. However much they question test reliability, they face a relentless media and political bombardment about test-defined success or failure which takes the reliability of national tests as a given. In short, the capacity of teachers to assess confidently and accurately is central to systemic improvement, and is depressed by a national testing system.

In the next section the virtues and difficulties of replacing tests by teacher assessment are discussed.

Teacher Assessment

Teacher assessment and the use of tests are not mutually exclusive. The current system involves both. Indeed, the present Key Stage 1 assessment method, which includes a large element of moderated teacher assessment, provides a basis for a way forward at Key Stage 2, although it would benefit from the relegation of the statutory test materials.

The principle and substantial advantage of teacher assessment over external testing is that it more easily integrates assessment with the curriculum and pedagogy. A single test may not cover the range of what has been taught which, in turn, is why today's teachers have had to develop advanced skills in teaching to the test. Narrowly focused testing may not take full account of the social and cultural backgrounds of all pupils and, thus, the ways their knowledge is framed.

The Role of Formative Assessment

The main benefit of assessment prior to the end of a Key Stage is to promote further learning. Formative assessment could be defined as any procedure designed for that purpose. It is worth repeating that with appropriate feedback on their work, pupils become more active and committed learners and their progress improves. Thus, the most effective assessments are built into the normal activities of lessons. Such assessments are based on both knowledge and intuition and are, therefore, informal. Requiring elaborate recording of such assessments is, therefore, pointless.

The objective of the current Assessment for Learning (AfL) Strategy is informed by this principle, although the linked Assessing Pupils' Progress (APP), which has a greater summative element, contains within it the potential for over-elaborate recording unless its use is under the control of teachers. Unfortunately, APP has the vice of appearing to teachers to be a centrally imposed requirement, which does not respect their professional practice. This issue is considered below.

Summative Assessment

Teacher summative assessment (TA), if conducted in a low stakes environment, can be at least as reliable and valid as the use of tests. If validity is the degree of correspondence between an assessment and what has actually been learnt, then TA may be more valid because it is more likely to cover the range of pupils' learning. The TA task is more likely to be expressed in the terms and context that pupils can understand; and it can be undertaken in non-threatening conditions[17].

Reliability

As argued above, testing is not as reliable as often assumed, and TA is more reliable than often assumed[18]. Research into TA confirms the capacity of teachers to assess the National Curriculum level achieved by their pupils. Evidence also suggests, however, that currently teachers are not always consistent in their own assessments or in comparison with other teachers.

The developers of tests constantly deal with gender, language and special educational needs (SEN) bias in test items and there is also a bias in TA. In particular, there is evidence of bias on ethnic

lines[19]. This is connected to evidence that teachers have a tendency to stereotype pupils according to ethnicity, but this bias can be overcome by two measures which will also have other important system benefits: moderation and professional development.

Moderation

Moderation of assessment between teachers at the same school and between teachers at different schools is a key aspect in developing reliability in TA. Moderation is any process by which teachers submit their judgements of pupil achievement to scrutiny. The evidence is that teachers appreciate the opportunity for peer discussion about their practice, both within a school and between schools. The fact that pupils are not identified within inter-school moderation enables questions of bias to be brought out and resolved neutrally. While all teachers improve their skills by means of such discussion and reflection, there is undoubtedly an opportunity to develop a cohort of teachers who are expert in assessment.

Whatever the approach to moderation, it must be teacher-led, locally organised, and be accompanied by a resource that supports national standardisation.

Continuing Professional Development (CPD)

In dealing with issues of standardisation and assessment bias as suggested in this paper, the Government should be contributing to a general improvement in the assessment skills of teachers.

The total resource applied to the current test system is substantial, if school staff time in preparing pupils is included as well as the £20 million annual cost of administering the tests themselves. If the same resource was directed instead towards developing teachers' capacity to make reliable summative assessments, including how to detect and eliminate bias[20], any amount of difficulties in TA would be resolved.

It has been a common aim amongst teacher organisations and support staff unions that the demand for, and supply of, appropriate professional development should be improved and that the introduction of a contractual right to CPD would stimulate rethinking at school level on its provision. The potential impact of integrating assessment into everyday pedagogy is so great that CPD in TA should be a national priority. However, the implementation of the AfL Strategy, and in particular APP does not provide a good model for this priority. Government should reflect on the way that good ideas and useful tools have been misused at school level because of the culture of compliance with perceived (in this case, mis-perceived) central imposition.

A one-off crash CPD programme would be unsuitable. Teacher assessment should be a permanent feature of the CPD offer for all teachers. As with much of the best CPD, local peer group discussion and reflection would be the most effective means and would integrate with inter-school moderation practice under the leadership of teachers expert in assessment.

Workload

Teachers suffer excessive workload which has often been imposed. Any proposals to change their practice are received in that context. It has been estimated that a Year 6 class teacher spends 400 hours in the year on end of Key Stage test preparation activities. It is difficult to make the equivalent estimate for a system of TA as a replacement for tests, but it is inconceivable that it could approach that figure. When test preparation is stripped away, the amount of time spent on assessment has the potential for reduction, not an increase.

For teachers, workload is not just about counting hours. They resent work which is imposed, unproductive and unnecessary. They give their time to do things which they believe contribute to their pupils' learning and over which they have professional control. For teachers, excessive

workload is about loss of control through centralised imposition in areas which should be in the locus of their proper professional judgement: curriculum detail, pedagogy, and assessment. When teachers are given the responsibility to make judgements on how to manage assessment in their classrooms in a non-bureaucratic environment, it will cease to be a workload issue because it will no longer be an imposition.

Summary of Proposals

The NUT and ATL welcome the decision of the coalition Government to "review how Key Stage 2 tests operate in future". Although both organisations believe that league tables are inappropriate mechanisms for school accountability, we also welcome the possibility of positive change through the coalition Government's decision to reform league tables to enable schools to focus on and demonstrate the progress of children of all abilities. Both organisations believe that the Government should formalise its position still further and initiate an independent review of the current National Curriculum assessment arrangements and the use of summative assessment for the purposes of institutional evaluation.

ATL and NUT jointly recognise that national assessment will be required at the end of the primary phase. The case is made above for the efficacy of teacher assessment for this purpose. Both organisations believe that, while teacher assessment outcomes should be reported to parents, they must not be used for league tables or any other public report for accountability purposes.

Both organisations are committed to annual sample testing in Year 6, on a national basis, covering the National Curriculum, in order for Government to re-establish an authoritative trend line of national performance.

Initial teacher education and CPD requires a radical overhaul in order to enhance the capacity and confidence of teachers to assess achievement and to embed formative assessment within everyday practice.

Moderation of teacher assessments, locally led by nationally accredited teacher experts, should be funded and supported. In order to support consistent moderation across England, there should be a national bank of assessment materials from which teachers can choose to draw to check their assessments.

Teachers need a restoration of properly accountable professional autonomy in both curriculum and assessment in order to empower them to make assessments.

References

[1] National Curriculum Task Group on Assessment and Testing (1988) *A Report 1988*, Department of Education and Science. London: HMSO.

[2] House of Commons – Children, Schools and Families Committee (2008) *Testing and Assessment Third Report of Session 2007 – 08, Volume I*, HC 169-I.

[3] Alexander R. (ed.) (2009) *Children, Their World, Their Education: Final Report and Recommendations of the Cambridge Primary Review.* Routledge.

[4] Harlen W. (2004) *A Systematic Review of the Evidence of the Impact on Students, Teachers and the Curriculum of the Process of Using Assessment by Teachers for Summative Purposes in Research Evidence in Education Library.* London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

[5] Harlen W. (2004) *A Systematic Review of the Evidence of the Impact on Students, Teachers and the Curriculum of the Process of Using Assessment by Teachers for Summative Purposes in Research Evidence in Education Library.* London: EPPI Centre, Social Science Research Unit, Institute of Education, University of London.

[6] Robinson C. and Fielding M. (2007), Primary Review Research Briefing 5/3 *Children and Their Primary Schools: Pupils' Voices.* University of Cambridge Esmee Fairbairn Foundation.

[7] Organisation for Economic Co-operation and Development (OECD) (2008) *Trends Shaping Education.* OECD.

[8] William D. (2001) *Level Best? Levels of Attainment in National Curriculum assessment*, ATL.

[9] De Waal A. (2008), *Fast Track to Slow Progress.* Civitas.

[10] BMRB Social Research Report Prepared for the General Teaching Council of England (2007) *Engaging With Parents: Pupil Assessment*. BMRB.

[11] Organisation for Economic Co-operation and Development (OECD) (2004). *Learning for Tomorrow's World: First Results from PISA 2003*, Executive Summary. France: OECD.

[12] Daugherty, R. et.al. (2004) *Learning Pathways Through Statutory Assessment: Key Stages 2 and 3. Final Report of the Daugherty Assessment Review Group.* Cardiff: Welsh Assembly Government.

[13] Organisation for Economic Co-operation and Development (OECD) (2007), *Reviews of National Policies for Education – Quality and Equity of Schooling in Scotland.* OECD.

[14] Tymms P. and Merrell C. (2007), Interim Report 4/1 *Standards and Quality in English Primary Schools Over Time.* University of Cambridge /Esmee Fairbairn Foundation.

[15] Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (2007), *Working Inside the Black Box: Assessment for Learning in the Classroom.* King's College London.

[16] 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 in Research Evidence in Education Library.* London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

[17] Gipps, C. and Stobart, G. (2003) *Alternative Assessment in T. Kellaghan and D . Stufflebeam (eds.) International Handbook of Educational Assessment.* Dordrecht: Kluwer Academic Publishers.

[18] Durant D. (2003), *A Comparative Analysis of Key Stage Tests and Teacher Assessments*, Paper Presented to British Educational Research Association Annual Conference, Edinburgh.

[19] Burgess S. and Greaves E. (2009), *Test Scores, Subjective Assessment and Stereotyping of Ethnic Minorities.* Centre for Market and Public Organisation, University of Bristol.

[20] Harlen W. (2004) *A Systematic Review of the Evidence of Reliability and Validity of Assessment by Teachers Used for Summative Purposes in Research Evidence in Education Library.* London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

the education union

www.atl.org.uk

www.teachers.org.uk

6875/06/10 RP