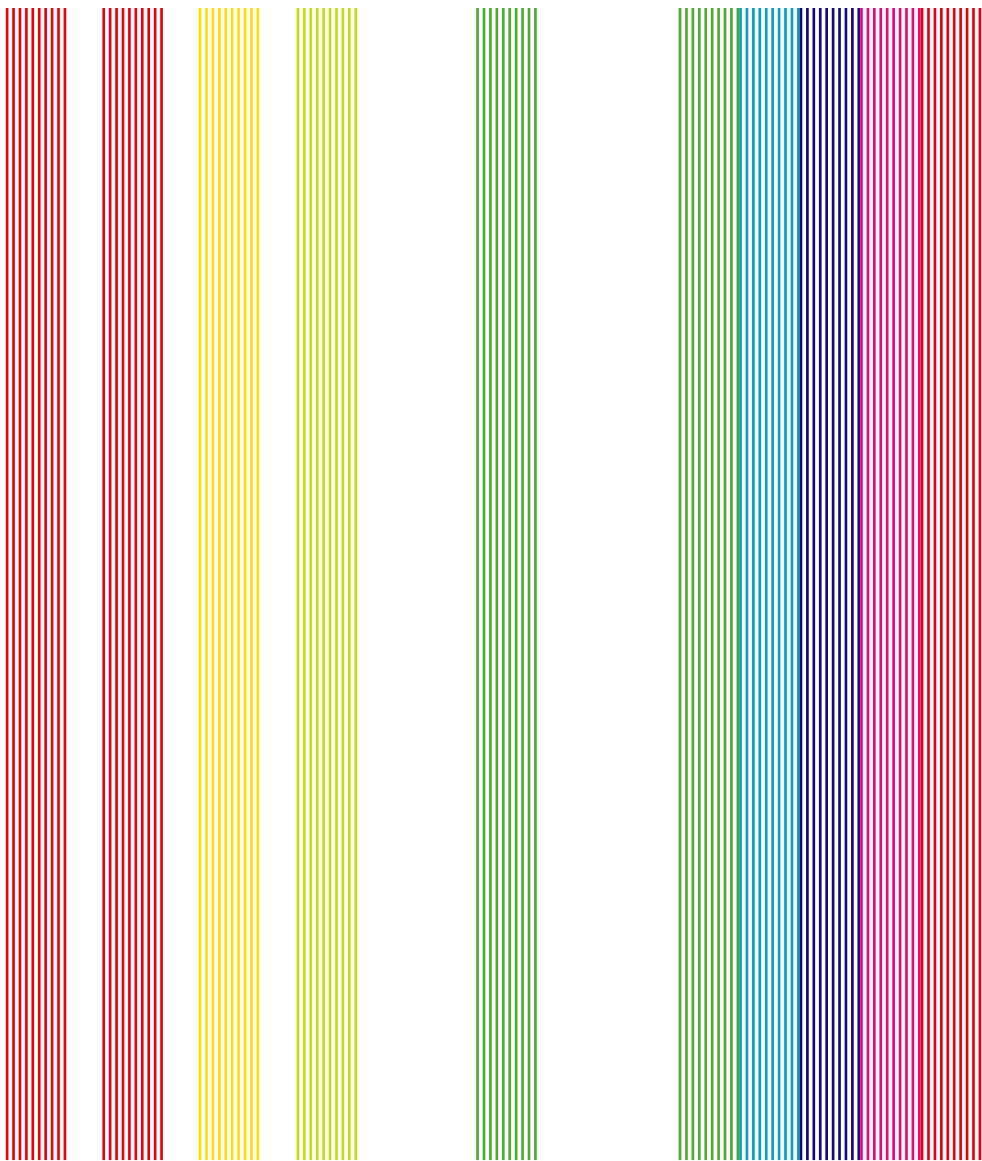


# The role of teachers in the assessment of learning

---



**ASSESSMENT REFORM GROUP**  
SUPPORTED BY THE NUFFIELD FOUNDATION

---

## The role of teachers in the assessment of learning

This pamphlet results from the *Assessment Systems for the Future* project, funded by the Nuffield Foundation. The project was set up by the Assessment Reform Group in September 2003 to consider evidence from research and practice about the summative assessment of school pupils, and to propose ways in which such assessment can benefit their education.

The role that assessment by teachers can take in summative assessment was the project's particular focus.

The project was directed by Wynne Harlen and based at the Faculty of Education, University of Cambridge. On the final page is a list of members of the project's Core Group, which has overseen all the projects activities and publications. Further information about the project and the Assessment Reform Group's other activities and publications can be found at [www.assessment-reform-group.org](http://www.assessment-reform-group.org)

## Summary

Assessment and testing have a strong effect on the lives and careers of young people. Decisions taken within and by schools influence the prospects and opportunities of their pupils and of even greater importance are their results of national tests and examinations. When the results of tests and examinations are used to pass judgments on teachers and schools, they also affect the ways in which pupils are taught. Given their importance, it is essential that results of summative assessment should reflect and influence school learning in the best possible way.

This pamphlet considers how to arrive at a comprehensive summative assessment system capable of providing information, based on sound evidence, about a wide range of pupil competences. Available research evidence leads to the conclusion that systems relying heavily on tests results are found wanting in several respects, particularly in their ability to give a dependable, that is, both valid and reliable, account of pupils' learning. It is argued that the negative consequences of summative assessment for learning and teaching can be minimised by more appropriate use of teachers' judgements.

At the same time it is acknowledged that a number of issues need to be addressed in implementing a system making use of teachers' assessment. Some key requirements are for: robust and permanent procedures for quality assurance and quality control of teachers' judgments; the provision of developmental criteria, which indicate a progression in learning related to particular goals; teachers to have access to well designed tasks assessing skills and understanding, which can help them to make judgments across the full range of learning goals; and for pre-service and in-service professional development that extends teachers' understanding and skills of assessment for different purposes. It is also important that summative assessment procedures are in harmony with the procedures of formative assessment and that they are transparent, with judgments supported by evidence so that all involved can have trust in the results.

Further, to avoid the negative consequences of using high stakes summative assessment to evaluate teachers and schools it is argued that systems of school accountability should not rely solely on the data derived from summative assessment of pupils and that the monitoring of standards of pupils' achievement should be derived from a wider base of evidence than test results from individual pupils.

Implications for those responsible for making assessment policy and for those responsible for implementing it are drawn out.

## Introduction

**A**SSESSMENT IS USED in many ways in education. A good deal of attention is now given to its use in helping teaching and learning, described as assessment *for* learning (AfL), or formative assessment. Here the focus is on assessment *of* learning, or summative assessment, which is used to summarise what pupils know or can do at certain times in order to report achievement and progress.

Since 2002 there has been a noticeable willingness in some assessment policy statements in the UK to consider alternatives to using external tests for summative assessment of pupils' performance. For example, the trend towards giving teachers a more central and professional role in summative assessment is shown in policy changes in England, in giving teachers of seven-year-olds more responsibility for pupil assessment from the summer of 2005, and in the reforms taking place in Wales, Scotland and Northern Ireland.

In Wales, testing seven-year-olds ceased in 2002 and the Welsh Assembly Government announced in the summer of 2004 that it would abolish national tests for 11 and 14-year-olds. In Scotland, teachers are drawing national assessment tasks from an electronic bank to support their judgments about pupils' attainment and test scores are no longer collected by the government. In Northern Ireland there are plans to end key stage assessment and the Transfer Test at 11+ and introduce a Pupil Profile based on teacher assessment. The profile will keep parents informed about the progress and

achievement of their children and will also inform crucial decisions such as transition to post-primary schools and choices at Key Stage 3.

In announcing the 2005 reform, the then schools minister for England, Stephen Twigg, said that he and his colleagues were "putting all our faith in teachers". He also said that trials conducted in 4,700 English schools had shown that teacher assessment of seven-year-olds was "robust", even though teachers are still required to use tests as part of the summative assessment process. The Chief Executive of England's Qualifications and Curriculum Authority also claimed that "We are, in this country, so much closer than our competitor countries to having the best possible basis for rigorous teacher assessment which... is based on far better evidence than at present, being gathered routinely and systematically by the children's teachers"<sup>1</sup>. He has predicted that external summative tests for 11-year olds and 14 year olds will eventually be replaced by moderated teacher assessment but that the transition could take 10 years. He envisages that teachers in England will one day be allowed to select tests for their pupils from a bank of assessment tasks and tests and choose when the tests should be taken.

The view taken in this pamphlet is that there are strong, evidence-based, reasons for taking forward this trend towards teachers having greater responsibility for summative assessment on a shorter timescale.

<sup>1</sup> Ken Boston, speech at the launch of the Annual Review of QCA, March 2005. ([http://www.qca.org.uk/251\\_12898.html](http://www.qca.org.uk/251_12898.html)):

Summative assessment by teachers can be problematic, and is certainly no panacea, but in many respects it is superior to an external test-based system.

The arguments presented are based on evidence from research, some of which is referenced here whilst the rest can be found in ARG publications<sup>2</sup>. The pamphlet begins by looking at some pros and cons of using teachers' assessment for summative purposes and the action needed to ensure high dependability (a combination of high reliability and validity) of the information it provides. It then reviews the problems associated with summative assessment systems based on tests and examinations and considers how to ensure that summative information fits the different uses to which it is put. The pamphlet concludes with implications for those involved in both policy and in the implementation of policy in education.

<sup>2</sup> Particularly in Gardner, J (ed) *Assessment and Learning* (2006) London: Sage.

Although the words 'pupils' and 'schools' are used here, the arguments and implications have relevance for learners in other educational institutions.

### Using teachers' assessment for summative purposes: pros and cons

**T**HERE ARE MANY different purposes for which pupils' work is assessed with a view to summarising their achievements. These vary from informal records of progress to high stakes certification (see Box 1) and occur in contexts across all phases of education from pre-school to adult learning. The ways in which assessment can be carried out also vary considerably. The concern here is to ensure that the way in which it is conducted provides information that is fit for its purpose.

The use of assessment by teachers for external summative purposes has long been advocated. The value of such a strategy

#### Box 1 The purposes of summative assessment in a national assessment system.

##### Individual pupils:

Uses *internal* to the school/college – for keeping records and giving reports on progress to other teachers, parents and pupils.

Uses *external* to the school/college – including certification, selection and meeting statutory requirements.

##### Groups of pupils:

*Evaluation* – of teachers, schools and local authorities. The types of evidence gathered are usually determined by national and local policy rather than by individual schools.

*Monitoring* – for year on year comparison of pupils' average achievements at the regional or system level. The procedures for doing this are also determined outside the school.

becomes particularly clear when one considers the qualities that effective summative assessment should have.

In common with assessment for other purposes, summative assessment should have the following qualities:

*Validity:* the assessment must cover all aspects, and only those aspects, of pupils' achievement relevant to a particular purpose.

*Reliability:* it should be designed so that users can have confidence that the results are sufficiently accurate and consistent for their purpose.

*Impact:* it should not only measure performance but have desirable consequences for teaching, learning and pupils' motivation for learning. Assessment generally has a strong impact on the curriculum and on pedagogy, so it is vital that any adverse effects are minimised.

*Practicability:* the resources required to provide it – teachers' time, expertise and cost, and pupils' learning time – should be commensurate with the value of the information for its users.

How successfully does summative assessment by teachers meet these four criteria? The table opposite weighs evidence for and against. However, before setting out the conflicting points it is important to be clear about the meaning given here to summative assessment by teachers. The definition that emerged from discussions the teachers and other professionals in education restricts the meaning to situations where teachers assess their own pupils:

*Summative assessment by teachers is the process by which teachers gather evidence in a planned and systematic way in order to draw inferences about their students' learning, based on their professional judgment, and to report at a particular time on their students' achievements.*

## Improving the reliability of teachers' summative assessment

IT IS FAIR to point out that the research identifying the deficiencies of teachers' assessment comes from studies where no attempt was made to prepare teachers for a major role in assessment. Other research<sup>4</sup>, and experience from countries where teachers' assessment is used for external summative purposes, identifies the conditions that affect the dependability of the assessment and the steps that can be taken to increase it.

The following five points were brought out in consulting those with experience of implementing summative assessment by teachers<sup>5</sup> in the UK and in Australia and the United States (Queensland and California). The project has also considered how pupils working towards national vocational qualifications are assessed in the workplace or college<sup>6</sup>.

1. Teachers should have clear criteria describing levels of progress in various aspects of achievement and, ideally, they

<sup>3</sup> See Appendix on costs of assessment in Working Paper 3 on ARG website.

<sup>4</sup> Summarised in Working Paper 2 (ASF, 2004) on ARG website.

<sup>5</sup> See report of seminars 2 and 3 and summary of the main points in Working Paper 1.

<sup>6</sup> See report of seminar 5 and Working Paper 4.

## Pros and cons of teachers' summative assessment

Pros	Cons
<b>Validity</b>	
<ul style="list-style-type: none"> <li>● Teachers can assess a wider range of achievement and learning outcomes than formal tests and examinations. Teachers' assessment can provide information about learning processes as well as outcomes.</li> <li>● Freedom from test anxiety and from practice in test-taking means that assessment by teachers gives a more valid indication of pupils' achievement.</li> </ul>	<ul style="list-style-type: none"> <li>● The validity of teachers' assessment depends on the learning activities and opportunities that schoolwork provides.</li> <li>● Bureaucratic moderation procedures for quality assurance could constrain the operation of teachers' summative assessment so that only "safe" and routine approaches are used.</li> </ul>
<b>Reliability</b>	
<ul style="list-style-type: none"> <li>● With appropriate training and moderation teachers' assessment can reach satisfactory levels of reliability.</li> <li>● There are many examples in other countries and in further and higher education of teachers making crucial summative assessments of pupils' performance.</li> </ul>	<ul style="list-style-type: none"> <li>● Teachers' assessment is often perceived as being, and indeed can be, unreliable and biased due to varying standards being applied.</li> <li>● Some external tests or tasks may still be needed to supplement teachers' judgments.</li> </ul>
<b>Impact</b>	
<ul style="list-style-type: none"> <li>● Teachers have greater freedom to pursue learning goals in ways that suit their pupils.</li> <li>● When teachers are gathering evidence from pupils' on-going work, information can be used formatively, to help learning, as well as for summative purposes.</li> <li>● Moderation procedures provide valuable professional development for teachers</li> <li>● Pupils can share in the process through self-assessment and derive a sense of progress towards "learning goals" as distinct from "performance goals".</li> </ul>	<ul style="list-style-type: none"> <li>● Public confidence in the system may be low due to teachers' assessment being perceived as inferior to external tests, particularly for children aged 11 and over.</li> </ul>
<b>Practicability</b>	
<ul style="list-style-type: none"> <li>● Financial resources are released at the school level by reducing the number of commercial tests purchased, which increased dramatically when national testing was introduced and used to judge schools.</li> <li>● Teachers can spend more time teaching rather than preparing for and marking tests.</li> <li>● Pupils' learning time is increased by at least two weeks per year by using classroom work rather than tests to assess progress<sup>3</sup>.</li> </ul>	<ul style="list-style-type: none"> <li>● Responsibility for summative assessment can increase the workload for schools and teachers.</li> <li>● Teachers can find that the process of moderating their judgments is time-consuming.</li> <li>● Training in the interpretation and use of assessment criteria is needed.</li> </ul>

should help to develop these criteria. As well as providing a common basis for interpretation of evidence, such criteria should also spell out the learning opportunities that are required. This makes it easier for teachers to assess pupils dependably on the basis of regular classroom work.

2. Professional development is needed so that teachers follow procedures that assure dependability. Training should also focus on the criteria to be applied and the sources of potential bias that have been revealed by research.
3. A system of moderation of teachers' judgments through professional collaboration benefits teaching and learning as well as assessment. Moderation that affects the planning and implementation of assessment, and consequently teachers' understanding of learning goals and of the criteria indicating progress towards them, has more than a quality assurance function.
4. The provision of a bank of well-designed tasks, with marking criteria, can do more than help teachers to make judgements about their pupils' achievements. Such tasks exemplify activities through which pupils can work towards important goals, such as critical reasoning and the application of knowledge in new situations. As assessment tasks they can provide pupils with interesting and relevant learning experiences. They should not be allowed to dominate the assessment process and certainly should not be seen as separate measures to be set beside teachers' judgments. Nor are they intended to confirm teachers' judgments. Rather they are part of the evidence that teachers can

use, if needed, to ensure that all intended goals are taken into account in their assessment.

5. Where teachers and users of summative assessment have become dependent upon external tests it will take time to increase teachers' competence in using on-going assessment and to build confidence in teachers' judgments. It is important that all involved have time to trial and evaluate new practices and to be clear about the procedures and the safeguards that are built in to protect dependability.

### Using tests for summative purposes: some problems

IT IS CLEAR that using teachers' assessment for summative purposes is not without its problems, some of which are shared by any procedure for summative assessment, particularly when the result is used for external high stakes purposes. However, these problems should be set against those caused by the alternative of depending on tests.

There is ample evidence that a system based on tests is flawed for the following reasons:

- It fails to provide information about the full range of educational outcomes that are needed in a world of rapid social and technological change and therefore does not encourage the development of these skills. These outcomes include higher-order thinking skills, the ability to adapt to changing circumstances, the understanding of how to learn, and the ability to work and learn collaboratively in groups as well as independently.

- It inhibits the development of formative assessment (or assessment for learning) which is proven to raise achievement levels and reduce the gap between higher and lower achieving pupils.
- The data it provides are less reliable than they are generally thought to be. For example it has been estimated that the key stage (KS) tests in England result in the wrong levels for at least a third of pupils at the end of KS2 and up to 40 per cent at the end of KS3<sup>7</sup>.
- The weak reliability of tests means that unfair and incorrect decisions will be made about some pupils, affecting their progress both within and between schools<sup>8</sup> and beyond school.
- There is no firm evidence to support the claims that testing boosts standards of achievement.
- It reduces some pupils' motivation for learning.
- It imposes stressful conditions that prevent some children from performing as well as they can.
- It encourages methods of teaching that promote shallow and superficial learning rather than deep conceptual understanding.

It is also evident that some test and examination results are being used for multiple purposes, including some for which they may only have limited value. The use of

individual pupil test results for a range of purposes, from target setting to league tables, is too simplistic. Information gathered for one purpose does not necessarily serve others, nor do the methods used to collect evidence of some types of learning suit all. The negative consequences of summative assessment may be minimised by giving teachers a greater role in assessing individual pupils<sup>9</sup> and using different approaches for evaluating and monitoring teacher and school performance, as discussed later.

Before considering the implications of this course of action, it is important to substantiate the claims that much summative assessment does not measure some key outcomes of modern education and that it is having a negative impact on pupils, on the curriculum, on teaching and on the use of assessment to help learning (AFL).

### *What is assessed?*

Much summative assessment – for example, the national tests that pupils in England take at the ages of 11 and 14 – depends on written tests of necessarily limited duration. As already noted, this format restricts the range of learning outcomes that can be assessed and excludes many of the higher-level cognitive and communication skills and the ability to learn both independently and collaboratively. The high stakes attached to the results encourage teaching to the test and excessive practising of test-taking. Research<sup>10</sup> confirms

<sup>7</sup> Wiliam, D. (2001) Reliability, validity, and all that jazz. *Education* 3–13, 29(3) 17–21.

<sup>8</sup> See Chapter 7 in J. Gardner (Ed) (note 1) and J. Gardner & P. Cowan (2005) *Assessment in Education* 12 (2) 145–164.

<sup>9</sup> See ASF Working Paper 2 available on the ARG website.

<sup>10</sup> See the summary in the ARG pamphlet *Testing, Motivation and Learning* on the ARG Website.

that this can result in pupils being taught to pass tests even when they do not have the skills that are supposedly being tested. A study commissioned by the Department for Education and Skills (DfES) concluded that while drilling 11-year-olds to pass national tests is likely to boost results it may not help pupils' longer-term learning<sup>11</sup>. The narrow range of learning outcomes assessed by tests contrasts with the broad view of learning goals reflected in the DfES *Every Child Matters* policy document<sup>12</sup>.

### *What ought to be assessed?*

It is crucial that assessment covers the learning that will be essential for young people who will live and work in a rapidly shrinking world and changing society. Two key sets of goals in any subject are:

- learning with understanding;
- and
- understanding learning.

The first refers to the development of “big ideas” – concepts that can be applied in different contexts, enabling learners to understand a wide range of phenomena by identifying the essential links between different situations. Merely memorising facts or a fixed set of procedures does not help young people to apply learning to a range of contexts.

The second set of goals relates to the development of awareness of the process of learning. It is widely recognised that

‘students cannot learn in school everything they will need to know in adult life’<sup>13</sup>. School must therefore provide the skills, understanding and desire needed for lifelong learning. Since what is assessed has a strong influence on what is taught and how it is taught, we must look critically at what is assessed. If the required outcomes are not included, then alternative methods of assessment are needed.

### *The impact on pupils, teachers and teaching*

As already emphasised, current testing regimes have considerable consequences for pupils' motivation and for their learning experiences. Research<sup>14</sup> indicates that

- test performance can become more highly valued than what is being learned;
- testing can reduce the self-esteem of lower-achieving pupils and can make it harder to convince them that they can succeed in other tasks;
- constant failure in practice tests demoralises some pupils and increases the gap between higher and lower achieving pupils;
- test anxiety affects girls more than boys;
- teaching methods may be restricted to what is necessary for passing tests (eg neglect of practical work).

These negative effects can operate both directly in the preparation of pupils for

<sup>11</sup> <http://www.dfes.gov.uk/research/programmeofresearch/projectinformation.cfm?projectId=14390&type=5&resultspage=21>.

<sup>12</sup> DfES 2004 see [www.everychildmatters.gov.uk](http://www.everychildmatters.gov.uk)

<sup>13</sup> OECD, 1999, *Measuring Student Knowledge and Skills*. OECD Programme for International Student Assessment (PISA). Paris: OECD, p9.

<sup>14</sup> See *Testing, Motivation and Learning* (2002), available on the ARG website.

external high-stakes tests, and indirectly in providing a poor model of assessment to teachers for use in all years of schooling

### *Assessment for learning*

It is also likely that opportunities to use assessment to help learning (and reduce the gap between higher and lower achieving children) are being missed. Worse, perhaps, is the distorting effect on assessment for learning (AfL), a process that the ARG defined as

*seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there*<sup>15</sup>.

Many schools give the impression of having implemented AfL when in reality the change in pedagogy that it requires has not taken place. This may happen, for example, when teachers feel constrained by external tests over which they have no control. As a result they are unlikely to give pupils a greater role in directing their learning, as is required in AfL, in order to develop the capacity to continue learning throughout life. The nature of classroom assessment is dictated by the tests. (see Box 2).

### **Box 2 How national tests have driven out assessment for learning.**

Longitudinal research in the PACE<sup>16</sup> project carried out in primary schools in England followed a cohort of pupils for eight years, beginning just before the introduction of national tests for seven-year-olds. It found that after the introduction of external tests, teachers' own classroom assessment became more summative. Prior to the introduction of tests, pupils felt that teachers' assessments helped their learning but they later noticed that their teachers increasingly focused on performance outcomes rather than learning processes. Pupils themselves began to adopt summative criteria in commenting on their own work.

### **Summative assessment that is fit for purpose**

#### *Reconciling formative and summative assessment*

Since the negative impact of tests on assessment for learning is one of the main reasons for proposing greater use of teachers' assessment it is important to consider how to reconcile the two.

Innumerable classroom events enable teachers to gather information about pupils by observing, questioning, listening to informal discussions and reviewing written work. In formative assessment this information may be used immediately to help pupils or it may be stored and used to plan future learning opportunities. The

<sup>15</sup> ARG (2002) *Assessment for Learning: 10 Principles*, available on the ARG website.

<sup>16</sup> Pollard, A., Triggs, P., Broadfoot, P., McNess, E and Osborne, M. (2000) *What Pupils Say: Changing Policy and Practice in Primary Education* London: Continuum.

information gathered in this way is often inconclusive and may be contradictory, for what pupils can do is known to be influenced by the particular context. This creates a problem for summative assessment but is useful for formative purposes. Serving the two purposes requires a distinction to be made between the *evidence* and the *interpretation* of the evidence.

For formative assessment the evidence is interpreted in relation to the progress of a pupil towards the goals of a particular section of work. Next steps are decided according to what has been achieved and what problems have been encountered. The interpretation is in terms of *what to do to help further learning*, not what level or grade a pupil has reached. For summative purposes, common criteria need to be applied and achievement is summarised in terms of levels or grades, which must have the same meaning for all pupils. This means that if the information already gathered and used formatively is to be used for summative assessment it must be reviewed against the broader criteria that define reporting levels or grades. Change over time can be taken into account so that preference is given to the best evidence that shows the pupil's achievement across a range of work during the period covered by the summative assessment.

Evidence of achievement can be used both to help learning and for reporting purposes, providing that the summative assessment is not a summary of formative judgments but a *re-evaluation* against the broader reporting criteria. Regular recording of grades and marks is not formative assessment, but a

series of 'mini-summative' assessments. It is, of course, necessary to have some quality assurance of the summative judgment. The more weight that is given to the summative judgment, the more stringent the quality assurance needs to be, preferably including between-school as well as within-school moderation of judgments.

Steps taken to increase the dependability of teachers' summative assessment for external purposes will inevitably improve their assessment for internal purposes. However, to avoid a negative impact on the formative use of assessment it is important that internal summative assessment is not carried out more often than is really required for reporting progress and achievement. Assessing pupils frequently in terms of levels or grades means that the feedback that they receive is predominantly judgmental, encouraging them to compare themselves with others. In such circumstances there is little attention by teachers or pupils to the formative use of assessment.

### *Assessment for evaluation*

Evaluation operates at several levels: the individual, the school, the local authority and the national. The impact of evaluation at each level depends on the type of information taken into account, the criteria used in judging effectiveness and the action that follows that judgment. When information is derived from summative assessment of pupils carried out for other purposes, it may not serve the purpose of evaluation. Some of the well-documented disadvantages of schools being judged by whether a specified percentage of pupils reaches a certain level are:

- the results are unlikely to reflect the full range of educational outcomes that a school strives for and for which it should be held accountable;
- disproportionate attention is paid to “borderline” pupils;
- it encourages focus on the narrow requirements of passing the test or examination.

Evaluation is best based on information about a range of pupil achievements and learning activities, judged by reference to the context and circumstances of the school. Schools should therefore provide information about the curriculum and teaching methods and relevant aspects of pupils’ backgrounds and learning histories. Some good examples exist in school self-evaluation guidelines<sup>17</sup>.

### *Assessment for system monitoring*

In the context of education, “monitoring” refers to changes in levels of pupil achievement, in provision or in teaching. It is usually associated with interest in whether “standards” are rising, falling, or remaining steady. Although the evidence used includes pupil achievement, the purpose is to inform policy and practice decisions, not to make judgments about individual pupils. Monitoring at school level is best undertaken within the context of self-evaluation, where other information needed

to interpret pupil assessment data is also collected. Even at the system level a change from one year to the next is unlikely to be meaningful; trends over longer periods provide more useful information.

The value of system monitoring depends on the range of information that is collected. The approach in England is to collect results from national tests, taken by every pupil in a cohort. For valid monitoring a wider range of evidence is needed, derived from observation of skills in action as well as assessment of products.

The economical advantage of collecting achievement data already available, as in using end-of-key-stage tests for identifying national trends, must be judged against the extent to which it provides useful and relevant information. Similarly, the cost of establishing and running surveys covering a wide range of educational outcomes has to be judged against the more detailed feedback that can be useful to both policy-makers and practitioners. Separating monitoring from the performance of individual pupils would obviate the need for central collection of individual pupil assessment data. This would remove the “need” for high stakes testing and would ensure that assessment – and, more importantly, what is taught – was no longer restricted to what can be tested. The continuation in several countries<sup>18</sup> of regular surveys of small random samples of pupils indicates the value of this approach.

<sup>17</sup> For example: *How Good is Our School* in Scotland, in England the emphasis on schools self-evaluation in *A New Relationship with Schools* (DfES and Ofsted, 2004) and in Wales *Guidance on the Inspection of Primary and Nursery Schools* (September 2004) and *Guidance on the Inspection of Secondary Schools* (September 2004) (Estyn).

<sup>18</sup> For instance, the National Assessment of Educational Progress (NAEP) in the United States, the National Education Monitoring Project (NEMP) in New Zealand and the Scottish Survey of Achievement (SSA).

## Conclusions

**T**HIS PAMPHLET has explained why there is concern that an assessment system mainly based on tests and examinations fall short of providing valid and reliable information about pupils' achievements. The reasons for this reside partly in the knowledge and abilities assessed, and partly in the methods of assessment and the use made of the results. The consequence is to constrain the curriculum and teaching methods and impair the implementation of assessment for learning. It also leads to summative assessment being carried out far too frequently. To minimise the unintended consequences, summative assessment should be designed to provide information for specific purposes and carried out only when progress needs to be summarised and evaluated. At other times teachers should focus on the formative use of assessment.

The evidence reviewed indicates that there are fewer negative consequences for learning and teaching in a system that makes more and, where it already exists, more appropriate use of summative assessment by teachers. However, it is acknowledged that this course of action would require a number of steps to be taken to meet the challenge of ensuring that teachers' summative assessment provides dependable information about pupils' achievements. Key points are:

- Robust and permanent procedures for quality assurance and quality control of teachers' judgments are needed to ensure that their summative assessment provides valid and reliable accounts of pupils' learning.
- Both pre-service and in-service professional development should extend teachers' understanding and skills of assessment for different purposes, highlight potential bias in teachers' assessment and help teachers to minimise the negative impact of assessment on pupils.
- Attention and resources must be given to creating developmental criteria, which indicate a progression in learning related to particular goals and can be applied to a range of relevant activities.
- Teachers should have access to well designed tasks assessing skills and understanding, which can help them to make judgments across the full range of learning goals.
- Procedures need to be transparent and judgments supported by evidence.
- Summative assessment must be in harmony with the procedures of formative assessment and should be designed to minimise the burden on teachers and pupils.

Further, to avoid the negative consequences of using high stakes summative assessment to evaluate teachers and schools:

- Systems of school accountability should not rely solely, or even mainly, on the data derived from summative assessment of pupils. Such data should be reported, and interpreted, in the context of the broad set of indicators of school effectiveness.
- The monitoring of standards of pupils' achievement should be derived from a wider base of evidence than test results from individual pupils. Teachers' assessment has a place in a system in

which a wide range of evidence is collected for small samples of pupils.

## Implications

THESE CONCLUSIONS have implications for those involved in framing policy at national and local levels, and for those who implement assessment policy in schools and colleges, for advisers and inspectors and for teacher educators. Whilst implementers are dependent on policy decisions, it is also the case that changes require understanding and appropriate action by those who must operate them in practice, particular, as here, where trust and responsibility are involved. There is, therefore, no hierarchy in the implications for policy-makers and policy implementers set out below.

### *Implications for national and local policy-makers*

- Recognise that the financial and time burdens at national and school levels of current summative assessment policies based on testing are not justified by the value of the information gained.
- Replace national testing, where it exists, by a requirement for reporting moderated teachers' judgments of pupil performance, and divert some of the time and money saved into quality assurance that enhances teaching and learning.
- Review the role of teacher assessment in examinations for 16 and 18-year-olds.
- Promote open discussion of why and how changes in the system are being made.
- Allow at least two years for the trial and evaluation of any new summative assessment system based on teachers'

judgments and a further similar period for dissemination to users and training of teachers'.

- Evaluate schools and encourage self-evaluation against a broad range of indicators, not only pupil achievement levels, thus reducing the negative impact of high stakes assessment.
- Set up a system of sampling pupils' performance for national monitoring, thereby reducing the overall test burden whilst increasing the breadth and relevance of the evidence.

### *Implications for school management*

- Establish a school policy for assessment that supports assessment for learning at all times and requires summative assessment only when necessary for checking and reporting progress.
- Arrange quality assurance of all summative assessment, including any tests given by teachers, so that decision made within a school about the progress of pupils are based on dependable information.
- Ensure that parents understand how assessment is helping learning and how criteria are used in reporting progress at given times during the year.
- Resist pressure for "hard" data from tests and encourage use of a range of types of evidence of pupils' learning.
- Provide protected time for quality assurance of teachers' assessment through moderation.

### *Implications for teachers*

- Ensure that assessment is always used to help learning and that, when a summative

assessment report is needed, the best evidence is reliably judged against relevant criteria.

- Involve pupils in self-assessment of on-going work and help them to understand the criteria used in assessing their work for reporting purposes and how summative judgments are made.
- Take part in moderation of summative judgments and other quality assurance procedures.
- Use tests only when most appropriate, not as routine.

### *Implications for inspectors and advisers*

- Review school policies and practices to ensure that assessment is being used formatively and is not overshadowed by summative tests and tasks.
- Encourage the use of a range of evidence of pupils' achievements.
- Ensure that continuing professional development in assessment is available for those who require it.
- Review the thoroughness of moderation and other procedures for quality assurance

and the extent to which they benefit teaching and learning.

- Help schools to develop action plans based on self-evaluation across a range of measures rather than only on levels achieved by pupils.

### *Implications for initial and professional development course providers*

- Ensure that courses allow adequate time for
  - discussion of the different purposes of assessment and the uses made of assessment data;
  - trainees and participants to identify, sample and evaluate different ways of gathering evidence of pupils' performance;
  - giving experience of generating assessment criteria linked to specific learning goals;
  - considering evidence of bias and other sources of error in assessment and how they can be minimised.

## Members of the Assessment Systems for the Future project Core Group

**Mr David Bartlett** Co-ordinator for Assessment, Birmingham City Council Children and Young Person Services

**Professor Paul Black** King's College, University of London

**Professor Richard Daugherty** University of Wales, Aberystwyth

**Dr Kathryn Ecclestone** University of Exeter

**Ms Janet English** Head Teacher, Malvern Way Infant and Nursery Schools, Croxley Green, Hertfordshire

**Professor John Gardner** Queen's University, Belfast

**Professor Wynne Harlen** University of Bristol, ASF project director

**Ms Carolyn Hutchinson** Head of Assessment Branch, Scottish Executive Education Department

**Professor Mary James** Institute of Education, University of London

**Mr Martin Montgomery** Assessment Development Manager, Council for the Curriculum, Examinations and Assessment, Northern Ireland

**Dr Paul Newton** Principal Assessment Researcher, Regulation and Standards Division, Qualifications and Curriculum Authority

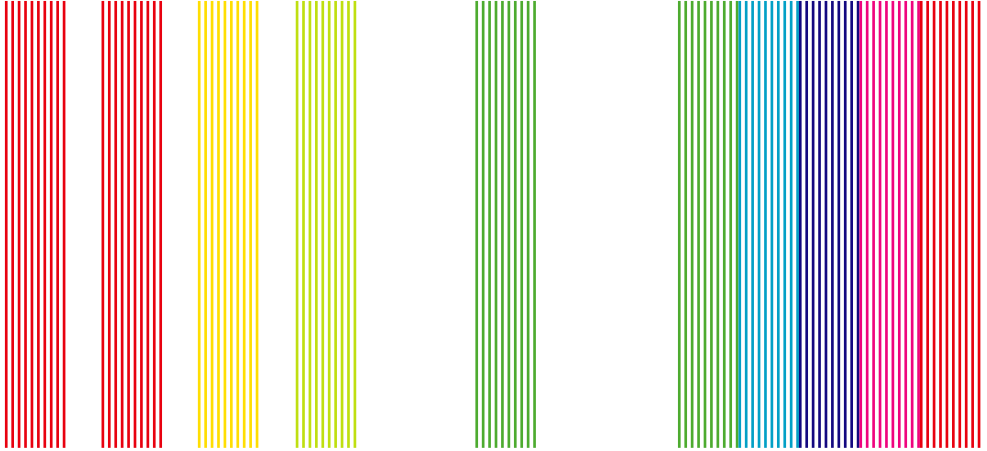
**Dr Catrin Roberts** Assistant Director, Nuffield Foundation

**Mr Jon Ryder** Teacher, Lord Williams's School, Thame, Oxfordshire

**Professor Judy Sebba** University of Sussex

**Dr Gordon Stobart** Institute of Education, University of London

**Ms Anne Whipp** Curriculum and Teacher Assessment Manager, ACCAC (the Qualifications, Curriculum and Assessment Authority for Wales).



**Further copies can be obtained from:**

CPA Office  
Institute of Education  
University of London  
20 Bedford Way  
London  
WC1H 0AL

**£3.00 including postage per single copy.  
Reduced rate for multiple copies.**

