



...children, their world, their education

## RESEARCH REPORTS

---

CPRT Research Survey I (new series)

# ASSESSMENT, STANDARDS AND QUALITY OF LEARNING IN PRIMARY EDUCATION

Wynne Harlen

For other reports in this series, for briefings on the reports and for reports and briefings in the first series (CPR) go to [www.cprtrust.org.uk](http://www.cprtrust.org.uk).

This report has been commissioned as evidence to the Cambridge Primary Review Trust. The analysis and opinions it contains are the author's own.

Copyright © Cambridge Primary Review Trust 2014

SUPPORTED BY

PEARSON



# **ASSESSMENT, STANDARDS AND QUALITY OF LEARNING IN PRIMARY EDUCATION**

**Wynne Harlen**

A report for the Cambridge Primary Review Trust

November 2014

This is one of a series of research reports commissioned by the Cambridge Primary Review Trust, a not-for-profit company established in December 2012 with the aim of consolidating and building on the evidence, findings and principles of the Cambridge Primary Review.

Cambridge Primary Review Trust is supported by Pearson Education, based at the University of York and chaired by Professor Robin Alexander.

A briefing which summarises key issues from this report is also available. The report and briefing are available electronically at the Trust's website: [www.cprtrust.org.uk](http://www.cprtrust.org.uk). The website also provides information about other reports in this series, and about the many publications of the Cambridge Primary Review.

We want this report to contribute to the debate about the future of primary education, so we would welcome readers' comments on anything it contains. Please write to: [administrator@cprtrust.org.uk](mailto:administrator@cprtrust.org.uk). The report contributes to the Trust's research review programme, which consists of specially-commissioned surveys of published research and other evidence relating to the Trust's eight priorities. This survey relates to priority 8, **assessment**:

Encourage approaches to assessment that enhance learning as well as test it, that support rather than distort the curriculum and that pursue standards and quality in all areas of learning, not just the core subjects.

**Professor Wynne Harlen** is former Director of the Scottish Council for Research in Education. Her earlier report, *The Quality of Learning: assessment alternatives for primary education* (2007) contributed to the research survey strand of the Cambridge Primary Review and in revised form was published in *The Cambridge Primary Review Research Surveys* (Routledge, 2010). Professor Harlen was also one of the contributing authors of *Children, their World, their Education: final report and recommendations of the Cambridge Primary Review* (Routledge, 2010).

Suggested citation: Harlen, W. (2014) *Assessment, Standards and Quality of Learning in Primary Education*. York: Cambridge Primary Review Trust.

Published November 2014 by Cambridge Primary Review Trust,  
Derwent College M, University of York, York, YO10 5DD, UK.

Copyright © 2014 Cambridge Primary Review Trust.  
All rights reserved.

The views expressed in this publication are those of the author. They do not necessarily reflect the opinions of Cambridge Primary Review Trust, Pearson Education or the University of York.

British Library Cataloguing in Publication Data:  
A catalogue record for this publication is available from the British Library.

**ISBN 978-0-9931032-0-9**

# ASSESSMENT, STANDARDS AND QUALITY OF LEARNING IN PRIMARY EDUCATION

## Introduction

Why and how we assess our pupils has an enormous impact on their educational experience and consequently on how and what they learn. This report provides a critical review of the assessment system in England introduced between 2014 and 2016, in the light of evidence from research and practice in six other countries. It begins with some ground-clearing discussion of the terms used in relation to tests and other forms of pupil assessment. The next two sections concern the purposes of assessment, particularly formative and summative assessment, the uses of summative assessment data for accountability and national monitoring and the impact on curriculum content and pedagogy. Section four describes how assessment for these purposes and uses is conducted in England, Scotland, Wales, Northern Ireland, New Zealand, Sweden and France, concluding with an overview of themes running through these examples. The main points from this analysis are drawn together in the fifth section, providing a critical perspective on the system in England in light of alternative approaches in other systems. Finally some implications for policy and practice are identified.

## 1 – CLARIFYING TERMS

Many of the terms used in discussing assessment have both technical and commonplace usage, which can sometimes impede good communication. While not everyone will agree on the definition of the terms used in assessment – and particular problems are associated with translation into other languages – it is important at least to make explicit the meanings being given to words used in this paper, in particular ‘assessment’, ‘testing’, ‘evaluation’ and associated terms such as standards, criteria, validity and reliability.

### Assessment, evaluation and appraisal

In its reviews of evaluation and assessment in its member countries, OECD (2013) makes a clear and useful distinction between pupil assessment, teacher appraisal, school evaluation and system evaluation, at the same time recognising that these system components need to work together in policies that aim to improve learning outcomes. For example:

The term ‘assessment’ is used to refer to judgements on individual pupil performance and achievement of learning goals. It covers classroom-based assessment as well as large-scale external tests and examinations. The term ‘appraisal’ is used to refer to judgements on the performance of school-level professionals, e.g. teachers and principals. Finally, the term ‘evaluation’ is used to refer to judgements on the effectiveness of schools, school systems and policies. (Nusche *et al* 2012: 24)

Assessment and evaluation both describe a process of generating and interpreting evidence for some purpose. They both involve decisions about what evidence to use, the generation and collection of that evidence in a systematic and planned way, the interpretation of the evidence to produce a judgement, and the communication and use of the judgement. It is worth noting that the evidence, of whatever kind, is only ever an indication or sample of a wider range that could be used.

### **Testing and other methods of assessment**

There is an important distinction between assessment and testing even though these terms are sometimes used interchangeably. Testing may be regarded as a method of collecting data for the process of assessment; thus assessment is a broader term, covering other methods of gathering and interpreting data besides testing. A closer look at what assessment involves helps to clarify this relationship and to identify other aspects of assessment involving terms such as 'standards' and 'criteria'.

All assessment of pupils' achievements involves the generation, interpretation, communication and use of data for some purpose. There are several forms that each of these components of assessment can take (Harlen 2013), which means that there is an enormous range of different kinds of activity that can be described as assessment. Each will involve: pupils being engaged in some activity; the collection of data from that activity by some agent; the judgement of the data by comparing them with some standard; and some means of describing and communicating that judgement.

For example, a standardised test comprises tasks created by an external agency that will have trialled the test during development using a large sample of the appropriate population, so that an individual's score can be expressed in terms of comparison with the average or 'norm' for that population. The result will indicate whether an individual pupil's performance is above or below average but not what he or she can do. A criterion-referenced test differs from a norm-referenced test by being designed to give information about what a pupil can do in relation to specified outcomes or standards. The items will be chosen for their relevance to these standards so that the results can be used in establishing not how a pupil compares with others, but what he or she can do in relation to the expected standard of skills and understanding. In theory, data collected by means other than by testing can also be judged either against the average performance or against criteria. However, in practice teachers' summative assessment – where teachers use a range of information from various sources to make judgements about whether certain standards of performance have been reached – is criterion-referenced.

### **Validity, reliability and dependability**

The *validity* of an assessment is generally defined in terms of how well what is assessed corresponds with the behaviour or learning outcomes that it is intended should be assessed. Various types of validity have been proposed depending on the kind of information used in judging the validity. For instance, *content validity* refers to how adequately the assessment covers the subject domain being taught and is usually based on the judgement of experts in the subject. However, content coverage is not enough to describe the full range of a test or

other assessment tool. *Construct validity* is a broader concept, reflecting the full range of outcomes of learning in a particular subject domain. The important requirement is that the assessment samples all aspects – but only those aspects – of pupils' achievement relevant to the particular purpose of the assessment. Including irrelevant aspects is as much a threat to validity as omitting relevant aspects.

Another form of validity, *consequential validity*, is not a property of the assessment instrument or procedure itself, but is a comment on how appropriate the assessment results are for the uses to which they are put. The validity of an assessment tool is reduced if inferences drawn on the basis of the results are not justified, either because things other than the intended construct are being assessed ('construct irrelevance') or because it does not include some aspects of the construct. For example, a test of arithmetic may be perfectly valid as a test of arithmetic but not valid if used to make judgements about mathematical ability more generally. This is described as 'construct under-representation'. The message is formally expressed in a widely quoted definition of validity by Messick:

Validity is an integrative evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment. (Messick 1989: 13)

The *reliability* of an assessment refers to the extent to which the results can be said to be of acceptable consistency and precision for a particular use. There are many factors that can reduce the reliability of an assessment. It is reduced if, for instance, the results are influenced by the particular set of items or tasks a pupil is required to attempt. It will be reduced also if the outcomes are dependent on who conducts the assessment (which teacher or oral examiner), or on who rates the pupils' assessment performances (which marker or observer). The particular occasion on which the assessment takes place, and the circumstances under which it is undertaken, can also affect the assessment outcome and contribute to reduced reliability. Thus reliability is often defined as, and measured by, the extent to which the assessment, if repeated, would give the same result.

Reliability has meaning mostly in the case of summative assessment and particularly for tests. When assessment is used formatively, as we see later, it involves only the pupils and the teachers; the notion of making a repeatable judgement and treating all pupils in the same way is not relevant. No judgement of grade or level is involved; only the judgement of how to help a pupil take the next steps in learning, so reliability in this formal sense is not an issue. However, high reliability *is* necessary when the results are used by others and when pupils are being compared or selected.

In relation to tests it would be misguided to assume that giving pupils all the same items and conditions provides them with equal opportunities. Individual pupils' responses to the same conditions will vary from day to day and, most importantly, the particular selection of items in a test will be found more difficult by some pupils than others of equal ability. All test items (apart from those in IQ-type tests) present questions in some context and there is research evidence that pupils who perform well in one item will not necessarily do so in another testing the same concepts or skills but set in a different context. The context effect

has been shown particularly in practical assessment, as in the science assessment of the APU (Qualter 1985) and by Pine *et al* 2006 in the US. But it also applies to written tests, since there is a large number of possible items of which only a small sample can be included in a test of a reasonable length. A different selection would produce a different result, giving rise to what is described as the 'sampling error'.

The sampling error can be much larger than is generally realised. For example, Wiliam (2001) estimated the difference that this would make for the end of key stages tests used in England in 2000. With a test of overall reliability of 0.80 this source of error would mean that about 32 per cent of pupils were likely to be assigned to the 'wrong' grade level, even though these levels each span roughly two years. A way of reducing this source of error would be to increase the number of contexts included for each competence assessed and thus the number of items used. But the length of a test cannot be greatly increased without incurring other forms of error (pupil fatigue, for instance) so more items per skills or concept would mean fewer skills and concepts included, thus reducing the range of what is assessed and so reducing the validity of the test. This illustrates a trade-off between validity and reliability which applies to whatever form the assessment takes, but it is most readily identified in relation to using tests. Attempts to increase reliability favour items assessing factual knowledge and the use of a closed item format that can be readily marked as correct or incorrect, limiting the range covered in the test. Attempts to increase validity by widening the range of items, say by including more open-response items where more judgement is needed in marking, will mean that the reliability is reduced.

The interdependence of the concepts of validity and reliability led to the introduction of the term *dependability* to refer to the 'intersection of reliability and validity' (Wiliam 1993). Although it has no meaningful numerical value, it can be a convenient term for the overall quality and usefulness of an assessment. In general, increasing the dependability of an assessment involves searching for optimum reliability whilst ensuring acceptable levels of validity, however these might be defined. This prioritises validity, but the trade-off has to take account of the intended use of the assessment results; some uses require greater efforts to improve reliability than others – which leads to matters of manageability (i.e. the cost in terms of time and resources of operating an assessment process of the required validity and reliability).

Stobart (2008) points out that there are several factors that 'pull against' the development of assessment procedures that provide the most dependable information for a particular purpose. One of these is 'the pull of manageability' which he describes as the search for simpler and more cost-effective assessments. This refers to resources of time, expertise and cost to those involved in creating and using the assessment materials and procedures. There is a limit to the time and expertise that can be used in developing and operating, for example, a highly reliable external test or examination. Triple marking of all test papers would clearly bring greater confidence in the results; observers visiting all candidates would increase the range of outcomes that can be assessed by external examiners; training all teachers to be expert assessors would have great advantages – but all of these are unrealistic in practice. Balancing costs and benefits raises issues of values as well as of technical possibilities.



## 2 – ASSESSMENT PURPOSES AND USES

It is generally agreed that there are two main purposes of pupil assessment:

- to help pupils while they are learning
- to find out what they have learned at a particular point in time.

These are described as *formative* and *summative* purposes of assessment.

Formative assessment has the purpose of assisting learning and for that reason is also called ‘assessment *for* learning’ (AfL). It involves processes of ‘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’ (ARG 2002).

Summative assessment has the purpose of summarising and reporting what has been learned at a particular time and for that reason is also called ‘assessment *of* learning’ (AoL). It involves processes of summing up by reviewing learning over a period of time, and/or checking-up by testing learning at a particular time.

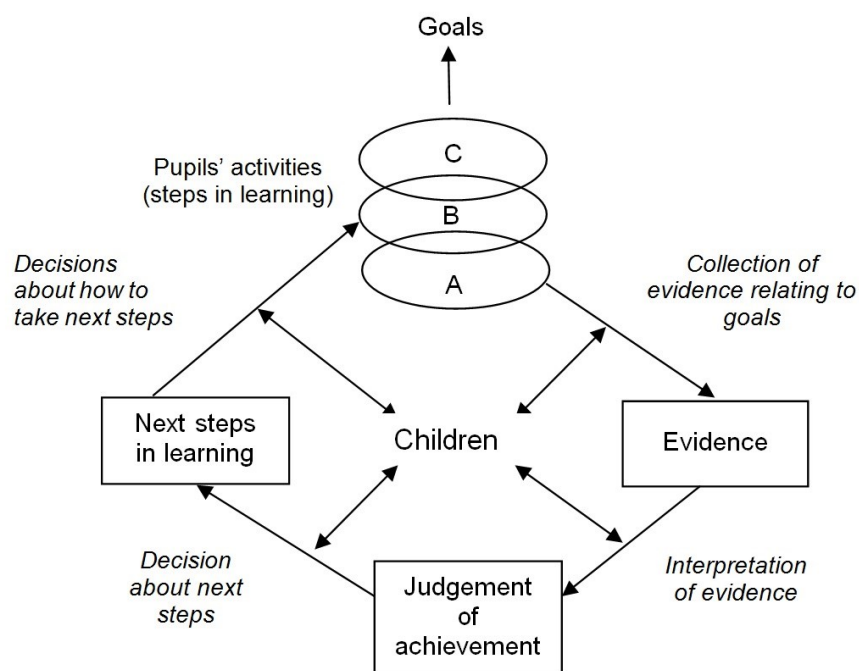
In theory it is possible to distinguish between *purposes* and *uses* of assessment, purposes being the reasons for conducting the assessment and uses being what is actually done with the assessment results (Harlen 2012). Newton (2007) points out, however, that whilst ‘assessment for learning’ is indeed a purpose, ‘assessment of learning’ is not a purpose but a statement of the subject of the assessment, which could have a number of purposes and uses. In the case of formative assessment there is only one main use of the information, which is to help learning. If it is not used for this purpose then the process cannot be described as formative assessment. By contrast, the data from summative assessment of pupils’ achievement (or assessment *of* learning) can be used in several ways, some relating to individual pupils and some to the aggregated results of groups or populations, not all of which are appropriate or valid uses (Newton 2012). Newton identified 16 uses of the results of national tests in England. These ranged from programme evaluation, target setting for pupils and schools, school monitoring and pupil selection to school choice by parents, and even the valuation of property in the areas of schools with high or low test scores. In some cases it is legitimate to use measures of pupil performance as part of the data used in making judgements (for instance, as an element in school evaluation), but use as a sole measure - and particularly where rewards and penalties are attached - it invites inappropriate actions to inflate the measured performance, a point to which we return later.

Before considering these two purposes of assessment in a little more detail, it is relevant to note that some authors on assessment have proposed the concept of ‘assessment *as* learning’ as a separate purpose of assessment. The main source of this idea is the work of Lorna Earl and her colleagues in Canada. They identify it with the role of pupils in their own assessment and the development of meta-cognition. However, separating pupils’ self-assessment from assessment for learning is not widely reflected in the current literature. Moreover, it has been suggested that conceiving assessment *as* learning relates to a less

positive relationship between assessment and learning in which assessment practices come to dominate learning experiences (Torrance 2007).

### Formative assessment

Formative assessment can be described as an on-going cyclic process (Figure 1) in which information is gathered in relation to the pupils' progress towards the short-term goals of a lesson or series of lessons. This information is then used to identify the appropriate next steps for the pupils and the action needed to take these steps. Pupils, of course, are the ones who do the learning so a key feature of formative assessment is the feedback that pupils receive about how to improve their understanding or skills or move on. At the same time the information gathered about pupils' progress provides feedback to the teacher, who can then adjust the pace or challenge of the learning activities – or regulate the teaching – to maximise opportunities for learning. Pupils, too, can have a role in decisions about their learning and direct their efforts more effectively if they know the purpose of their activities. This means not just knowing what to do but what they are trying to achieve in terms of quality as well as goals.



**Figure 1: Assessment for formative purposes (adapted from Harlen 2006)**

In summary, the features of formative assessment are:

- Pupils are engaged in expressing and communicating their understandings and skills through classroom dialogue, initiated by open and person-centred questions.
- Pupils understand the goals of their work and have a grasp of what is good quality work.
- Pupils are involved in self-assessment so that they take part in identifying what they need to do to improve or move forward.

- Teachers give feedback to pupils that provides advice on how to improve or move forward and avoids comparisons with other pupils.
- Teachers use information about ongoing learning to adjust teaching so that all pupils have the opportunity to learn.
- Dialogue between teacher and pupils encourages reflection on their learning.

The reason for attention to formative assessment lies in the evidence of its effectiveness in improving learning. Empirical studies of classroom assessment have been the subject of several research reviews. The review by Black and Wiliam (1998a) attracted attention worldwide partly because of the attempt to quantify the impact of using formative assessment. A key finding was that 'improved formative assessment helps the (so-called) low attainers more than the rest, and so reduces the spread of attainment whilst raising it overall' (Black and Wiliam 1998b). Since then there have been a number of other reviews and investigations which have justified the considerable claim made for improved pupil learning.

### **Summative assessment**

Summative assessment has the purpose of summarising and reporting what has been learned at a particular time and for that reason is also called 'assessment of learning' (AoL). It involves processes of summing up by reviewing learning over a period of time, and/or checking-up by testing learning at a particular time.

Since formative assessment is defined as helping learning, there is tendency to regard it as the 'good' face of assessment and summative assessment as the reverse. But summative assessment is important for a number of reasons. First, whilst it is not intended to have direct impact on learning as it takes place, as does formative assessment, it nevertheless can be used to help learning in a less direct but necessary way as, for example, in providing a summary of pupils' learning to inform their next teacher when pupils move from one class or school to another. Second, it enables teachers, parents and schools to keep track of pupils' learning, both as individuals and as members of certain groups (such as those who are high achievers and those who need special help). Third, it provides data which, together with contextual factors, can be used for school evaluation and improvement. Finally, it cannot be avoided: teachers have little choice about whether and when they conduct summative assessment since requirements and procedures are generally established at school or national level, not by individual teachers. By contrast, formative assessment could be considered, in a sense, to be voluntary in that it is possible to teach without it and it is part of the process of teaching which teachers largely decide for themselves. Formative assessment can be urged in official documents but cannot be mandated in the way that summative assessment can be required by statute.

The key features of summative assessment are that it:

- relates to achievement of broad goals expressed in general terms rather than the goals of particular learning activities;
- its results are reported at certain times, not as an everyday part of learning;
- uses evidence obtained when pupils are involved in special tasks or tests as part of, or in addition to, regular work;

- may be based on teachers' judgements, tests or a combination of these;
- involves judging the achievement of all pupils against the same criteria or mark scheme;
- requires some measures to assure reliability;
- typically provides limited, if any, opportunities for pupil self-assessment.

For individual pupils, the uses of summative assessment can be described as either 'internal' or 'external' to the school community:

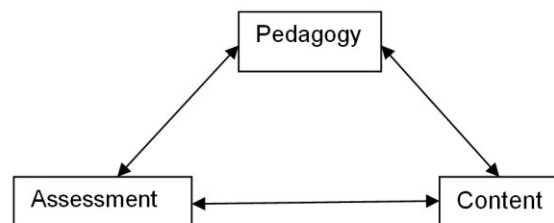
- 'Internal' uses include using regular grading, record keeping and reporting to parents and to the pupils themselves; at secondary level, informing decisions about courses to follow where there are options within the school.
- 'External' uses include meeting the requirements of statutory national assessment, for selection, where selective secondary schools exist; at the secondary level, certification by examination bodies or for vocational qualifications, selection for further or higher education.

In addition to these uses relating to judgements about individual pupils, results aggregated for groups of pupils are used for evaluating the effectiveness of the education provided by teachers, schools and local authorities. The main uses of aggregated results in England are:

- Accountability: for evaluation of teachers, schools, local authorities
- Monitoring: to compare results for pupils of certain ages and stages, year on year, to identify change in 'standards'.

### 3 – THE IMPACT OF ASSESSMENT

The relationships among assessment, pedagogy and content are frequently expressed as triangular, but often with 'curriculum' at one apex. However, if we take 'the curriculum' to mean all that is experienced by pupils at school, then each of these is an aspect of the experienced curriculum, as in Figure 2. Here we refer to 'curriculum content' to make clear that content is just one aspect of the curriculum experienced by pupils, while the whole includes content, pedagogy and assessment and the underlying assumptions about learning.



**Figure 2: Interactions among aspects of the whole curriculum**

In Figure 2, the arrows acknowledge what is well known – that what we teach is influenced by how we teach, and what and how we assess influences both how and what we teach.

These interactions are important, since it is no use advocating the use of more pupil-oriented or inquiry-based teaching if there is an overbearing assessment (whether by testing or teachers' judgements) or a curriculum overcrowded with content. It is no use suggesting that the content should be focused on conceptual understanding ideas if the assessment requires memorising multiple facts or if the pedagogy does not forge links that are necessary to form these overarching ideas. It is no use wanting pupils to develop responsibility for their own continued learning if teaching does not allow time for reflection and room for creativity.

But the impact of assessment on the curriculum content and teaching approach is by no means necessarily a negative one. An effective assessment system supports learning in a variety of ways, from providing formative feedback for use in short-term decisions about learning activities to providing information about pupils' achievement for reporting to parents, for use in longer-term planning and as part of school self-evaluation. Through establishing criteria for achievement or providing tasks that exemplify the use of inquiry skills and understanding, assessment can help to clarify and communicate the meaning of learning objectives.

Negative impacts arise when what is assessed reflects only easily tested aspects of learning, compounded by attaching rewards and punishments to the results, acquiring 'high stakes'. The pressure on teachers to increase test results is transferred to pupils, even if the tests are not high stakes for pupils. Research shows that when this happens, teachers focus teaching on the test content, train pupils in how to pass tests and feel impelled to adopt teaching styles which do not match what is needed to develop real understanding. Initially this effort increases test scores, as reported for the national tests in England by Wyse *et al* (2010) but soon level off as the effect degrades. Then the results become meaningless in terms of intended learning.

There is now a large body of research evidence on the negative impact of high stakes use of performance measures not just in education but in other social services, too, where naïve indicators are used as targets. The effect is to occlude the value of the indicators as a measure of quality, neatly expressed as Goodhart's Law: 'when a measure becomes a target it ceases to be a good measure' (see Elton 2004/5).

Whilst those responsible for accountability policies that lead to these impacts may protest that this was not the intention – that 'teachers do not need to teach to the tests' – and may even agree that some of the effects are undesirable, yet the practice continues. The mechanism for this influence is to set targets for schools, districts, or the nation as a whole to raise their test results. One may well suspect that holding teachers accountable for raising test scores is a deliberate use of the impact of assessment on curriculum content and a direct way of controlling what is taught. As Stobart (2008) comments, this is a quicker and cheaper way of changing curriculum content and pedagogy than developing curricula and providing professional development.

## 4 - ASSESSMENT SYSTEMS: ENGLAND AND SOME OTHER COUNTRIES

In this section we review briefly some key aspects of the assessment system at the primary level in England and, for comparison, in the other countries of the UK and in New Zealand, Sweden and France. In each case we consider how the system provides for formative and summative use of individual pupil assessment, for accountability and for the national monitoring of standards. To provide the context for the description of the assessment procedures we give in each case some background information concerning the system and primary curriculum.

### England

#### *Background*

In England (population 53 million) education is compulsory from the age of 5 to 17. The legal requirement is that pupils must be in some form of education or training to the end of the academic year in which they reach the age of 17. From 2015 this becomes the age of 18. Most children attend some form of pre-school education or care before the age of 5 and there are legal requirements to be met by early years providers which include Early Learning Goals.

About 93 per cent of pupils are in state funded schools of varying kinds. Others are in private (independent) mostly secondary, schools, which charge fees, are not funded by the government and do not have to follow the national curriculum.

The national curriculum for the compulsory years is structured in four key stages: Key Stage 1 (ages 5-7); Key Stage 2 (ages 7-11); Key Stage 3 (ages 11-14); Key Stage 4 (ages 14-17). There is also a statutory framework concerning the care and experiences of younger children attending any form of pre-school provision up to the age of five. The framework for the Early Years Foundation Stage identifies seven areas of learning – three ‘prime’ areas and four ‘specific’ areas – each with explanatory notes and details.

#### *Primary curriculum and assessment*

A revised National Curriculum for England, introduced in 2014, sets out the programmes of study for the subjects to be taught. For the two key stages of the primary school the subjects are:

- Key Stage 1 (ages 5-7): English, mathematics, science, art and design, computing, design and technology, history, geography, music, physical education, religious education.
- Key stage 2 (ages 7-11) as for key stage 1 with the addition of foreign language.

For the core subjects of English, mathematics and science the curriculum is set out in considerable detail, with non-statutory expansions and examples. For these core subjects, KS2 is divided into ‘lower KS2’ (Years 3 and 4) and ‘upper KS2’ (Years 5 and 6). A general

statement of what is intended in KS1 and in each of the two halves of KS2 is followed by the programme to be taught in each year. For the non-core subjects KS2 is not divided and the programme of study is expressed only in general terms for the whole of KS1 and KS2.

There is a marked difference in the way the curriculum is expressed between the new curriculum and the previous (1999) version. No longer are there 'attainment targets' setting out in levels of increasing difficulty the knowledge and skills pupils are expected to have. In the 1999 curriculum eight levels, plus a description of exceptional performance, were used in reporting pupils' attainment. In the 2014 curriculum the attainment targets are identified as what 'pupils are expected to know, apply and understand' as specified in the various programmes of study. The abandonment of levels was recommended by the expert panel set up to consider the curriculum revision (DfE 2011). Among the reasons for this was lack of information about what achieving a level means in terms of what pupils are able to do and of any indication of what is necessary to make progress. Other concerns related to 'labelling' of and by pupils, the use of levels as targets and the impact on pupils regarded as unable to reach target levels resulting in reduced opportunities for progress. Also, as the levels were well spaced (about two years apart), the practice had developed of creating sub-levels of questionable meaning in terms of cognitive development and serving only to prescribe the curriculum more closely.

To replace levels as a means of reporting pupils' performance at the end of key stages, new 'performance descriptors', designed to inform teachers' assessment, have been created. After consultation on drafts, final versions are to be published in 2015 for implementation in 2016. Performance descriptors in the core subjects identify what pupils are able to do if they meet certain standards. For KS1 teachers assess pupils in mathematics, reading and writing as meeting one of four standards: mastery standard; national standard; working towards national standard; below national standard. The descriptor for each standard comprises a number of statements of what pupils should be able to do in each aspect of the subject, the total comprising 50 or so such elements. Using professional judgement and evidence from class work and a range of activities, teachers decide which standard most closely matches a pupil's overall attainment. Pupils must demonstrate a majority of the elements described for working at a standard. For science at KS1 and KS2 there is a single descriptor ('working at national standard'). At KS2 for writing, which is only assessed by teachers' assessment, there are five descriptors but for mathematics and reading, for which there are end of KS2 national curriculum tests, there is only one in each case.

### *Purposes and methods of assessment*

The formative use of assessment continues to be encouraged but without any specific support in terms of funding or government guidelines. The summative assessment pupil performance is explicitly described as having two aims: 'to provide standard information to parents and to give a picture of school performance' (DfE 2014: 6).

For the first of these aims, assessment begins early. For children aged 2 and 3 in early years settings, practitioners must review and write a short summary of each child's development in the 'prime' areas of learning of the Early Years Foundation Stage (EYFS) (DfE 2012). Reporting on the 'specific' areas is optional. The EYFS profile, used to assess a child in the

years in which they reach the age of five, is no longer compulsory since the introduction of a new 'reception baseline' (DfE 2014). However, the EYFS curriculum and goals remain statutory.

The reception baseline to be used from 2016 is an assessment of all pupils in the first few weeks of starting school using a measure chosen by the school from a government list of published tools meeting required criteria. Although the results, supplemented by teachers' observations, will be reported to parents, the main use of this baseline is for accountability. The content of the baseline measure has to reflect the communication and language, literacy and mathematics areas of the EYFS and also demonstrate 'clear progression towards the key stage 1 national curriculum in English and mathematics' (STA 2014: 1).

The progress made by pupils from the reception baseline to the KS2 tests (the 'progress standard') will be used as a basis for judging the performance of all-through primary schools. For schools with only KS2 the progress from KS1 to KS2 tests will be the measure. In both cases, an alternative measure is the percentage of pupils attaining the expected standard at KS2 (the 'attainment standard'). Schools will be considered to be above the 'floor standard' if they meet either the standards of progress or of attainment (85 per cent or more achieving the new expected standard). Both of these measures refer to attainment in reading, writing and mathematics.

Teacher assessment is the main method for assessment at the end of KS1, informed by tests in mathematics and reading externally set but marked by teachers. From autumn 2014 statutory teacher assessment will be in terms of new performance descriptors and from 2016 test results will be expressed as a scaled score (a score which has been translated from a raw score onto a score on a fixed, defined scale). Teacher assessment of speaking and listening and of science will continue to be reported.

At the end of KS2 teacher assessment in mathematics, science, reading and writing is reported. Pupils also sit tests in mathematics, reading, grammar, punctuation and spelling, which are externally set and marked. As for KS1, results of tests will be reported to pupils and parents as scaled scores from 2016 and parents will be given information about their child's score together with the average for the school, the local area and nationally.

For non-core subjects schools are free to use their own assessment and reporting methods. After a competition, nine schools have been given a small fund to develop new assessment methods designed to replace the system based on levels and to make these available to other schools.

### *National monitoring*

For the core subjects where there are end of KS2 tests, aggregation of scaled scores in each subject provides the picture of national performance. As there are no KS2 tests for science, the national performance is provided by an annual test of a sample of pupils. Plans are for this to continue and to be extended to include practical assessment.



## Scotland

### *Background*

In Scotland (population 5.3 million) education is compulsory from the age of 5 to 16. Children attend primary school for seven years and transfer to non-selective secondary schools at the age of 11/12. Neither the curriculum nor its assessment is governed by legislation in Scotland, as it is in the rest of the UK. In the absence of regulation, factors which ensure implementation of changes include a tradition of conforming to central policy and wide consultation on changes. Inevitably, achieving consensus is a slow process and often means that change is evolutionary. Thus the development of the curriculum guideline, *Curriculum for Excellence*, in place at the time of writing, was begun with a national debate on education in 2002 and finally implemented in schools from 2011.

At the core of *Curriculum for Excellence* are the ‘experiences and outcomes’ which describe for each subject area the expectations for learning and progression. These are expressed at five levels:

- Early – the pre-school years and the first year of primary school (P1)
- First – to the end of the fourth year (P4)
- Second – to the end of the primary years (P7)
- Third – the first three years of secondary school (S1 to S3)
- Fourth – the upper years of secondary school or college.

The expectations at the end of these periods for each theme within each subject are expressed as ‘I can’ statements which combine the content and process of learning. For example, the second level statement for ‘estimating and rounding’ in the mathematics curriculum states ‘*I can use my knowledge of rounding to routinely estimate the answer to a problem then, after calculating, decide if my answer is reasonable, sharing my solution with others*’.

Teachers are responsible for deciding what to teach each year in order to meet the expectations set for the end of each period.

In Gaelic medium education the Scottish curriculum is delivered through the medium of Gaelic. It is a Gaelic language approach to learning and it provides cultural opportunities and experiences while endorsing the principles of Curriculum for Excellence.

### *Primary curriculum and assessment*

The eight curriculum areas of primary education are, in alphabetical order: expressive arts; health and wellbeing; languages; mathematics; religious and moral education; sciences; social studies; technologies. There are no subjects designated as ‘core’ but special attention is given to literacy and numeracy, which are the subject of national monitoring.

### ***Purposes and methods of assessment***

The framework for assessment in *Curriculum for Excellence* built on the strengths of effective approaches to assessment in Scotland developed through the programme *Assessment is for Learning* introduced in 2003 (SEED 2004). The new framework preserves the emphasis on formative use of assessment, identifying the purposes of assessment as:

- to support learning that develops the knowledge and understanding, skills, attributes and capabilities which contribute to the four capacities (successful learners, confident individuals, responsible citizens, effective contributors)
- to give assurance to parents, to learners themselves, and others, that children and young people are progressing in their learning and developing in line with expectations
- to provide a summary of what learners have achieved, including through qualifications and awards
- to contribute to planning the next stages of learning and to help learners progress to further education, higher education and employment
- to inform future improvements in learning and teaching (Scottish Government 2011)

Teachers assess pupils' performance against the standards and expectations set out in the experience and outcomes of *Curriculum for Excellence* using information from pupils' involvement in a wide range of tasks and activities including dialogue and interactions with peers and teachers, practical investigations, performances, reports, as well as specific assessment tasks and activities. There is emphasis across the curriculum on assessment in literacy and numeracy, health and wellbeing, ICT and higher order skills including creativity.

To help teachers with their assessment teachers have on-line access to a comprehensive National Assessment Resource (NAR), a single place in which assessment materials for *Curriculum for Excellence* can be stored and shared. It includes assessment materials developed by the Scottish Qualifications Authority, Education Scotland and practitioners. It provides exemplars and materials showing how practitioners have created opportunities for on-going assessment integrating assessment and learning, shows how standards and expectations are interpreted in practice and provides examples that can be a focus for moderation. Experienced teachers can add their own resources to the NAR.

At the end of P7, when pupils are transferring to secondary school a summative assessment of their achievements is provided in the form of a pupil profile. This draws together information about the pupil's learning in a form developed by each school. A typical profile provides a judgement as to whether progress within the level is 'developing', 'consolidated' or 'secure' for each subject area and level of the *Curriculum for Excellence*.

### ***National monitoring***

The Scottish Survey of Literacy and Numeracy (SSLN) is a national sample-based survey which monitors performance in literacy and numeracy in alternate years at P4, P7 and S2. The SSLN is fully aligned with *Curriculum for Excellence* and samples learners' achievements

to measure standards over time and to monitor national performance in literacy and numeracy. Pupils can take the tests in English or Gaelic. Teachers of P4 or P7 classes are asked to complete an online teacher questionnaire.

Findings from the surveys are used to inform the content of the NAR and professional learning resources for practitioners to facilitate improvements in learning, teaching and assessment at classroom level. Results of the SSLN are published each year in the spring following survey participation. Individual pupil results are not available since different pupils attempt different questions. Results for all the assessments are put together to build up a picture of overall performance across Scotland. Results are not available at a school or local authority level.

## **Wales**

### ***Background***

Education in Wales (population 3 million) is compulsory from age 5 to 16, although most attend some form of school from the age of 3 to 18. There are different types of state schools in Wales as in England, but no academies run directly by the government. State secondary schools in Wales are fully comprehensive and there are no grammar schools. About a quarter of primary and a fifth of secondary pupils are taught in Welsh-medium schools or in Welsh-medium classes. All pupils are taught the Welsh language throughout their school education.

Soon after responsibility for education was devolved from Westminster to the Welsh Assembly Government (now the Welsh Government), the Department for Children, Education and Lifelong Learning (DCELLS) developed a new curriculum covering the age range three to 18 and introduced in 2008.

The school years are divided into:

- Foundation Phase (pre-school years plus key stage 1) (ages 3 to 7)
- Key stage 2 (ages 7 to 11)
- Key stage 3 (ages 11-14)
- Key stage 4 (ages 14-16)

A non-statutory skills framework across the full age range 3 to 18 gives guidance about continuity and progression in developing thinking, communication, ICT and number skills.

### ***Primary curriculum and assessment***

In the Foundation Phase (which combines pre-school and KS1) the skills are developed across seven statutory areas of learning:

- personal and social development, well-being and cultural diversity
- language, literacy and communication skills (English or Welsh)
- mathematical development

- Welsh language development
- knowledge and understanding of the world
- physical development
- creative development.

For each area the *Foundation Phase Framework* (WAG 2008) sets out the opportunities to be provided to children in a series of bullet points indicating the skills to be developed and the range of contexts for using the skills. The same document provides Foundation Phase Outcomes describing the type and range of achievements characteristic of children within the Foundation Phase. There are six descriptions of outcomes at progressive levels of development. In deciding a child's outcome at the end of the Foundation Phase, teachers judge which description best fits the child's performance. The upper four outcome descriptions link to national curriculum levels.

At the end of the Foundation Phase, teachers are required to assess and report outcomes attained by each child by means of teacher assessment in:

- Personal and social development, well-being and cultural diversity
- Language, literacy and communication skills in English or Welsh
- Mathematical development.

The KS 2 curriculum retains the structure of a previous national curriculum which applied in England and Wales in 1999. It specifies requirements for separate subjects: Welsh, English, mathematics, music, science, ICT, design and technology, history, geography, art and design and physical education. Modern foreign language is added at KS3. English, Welsh, Welsh second language, mathematics and science are identified as core subjects. For each subject programmes of study specify what pupils should be taught and attainment targets set out the expected standards of pupils' performance in eight level descriptions of increasing difficulty, with an additional description above Level 8 for exceptional performance. At the end of KS2 it is expected that the great majority of pupils should be within the range of Levels 2 to 5.

### ***Purposes and methods of assessment***

The formative use of assessment is given prominence in the guidance *Making the Most of Assessment 7-14* (WAG 2010). Teachers are urged that assessment for learning should drive day-to-day assessment and be aligned with assessment of learning.

For summative assessment and reporting there are statutory requirements for annual reporting to parents and for end of KS2 reporting of attainment against national levels. Together with more detailed teacher assessment information about individual learners, they also form the main evidence base for receiving schools at transition. Cluster group moderation of end of KS2 teachers' judgements is required in the core subjects. These results contribute to annual national data collection.

It is a statutory requirement for all maintained schools in Wales to use Welsh Government reading and numeracy tests every year for learners in Years 2 to 9, or who are 7 to 14-years-

old. The tests are available in English and Welsh. These are written tests, administered and marked in the school. All the tests are available in English but pupils in years 2 and 3 in Welsh-medium schools take the reading test in Welsh. No other tests are required. Teacher assessment in all non-core subjects for each pupil each year are collected by the Welsh Government and reported in an annual report which looks at performance by gender, subject and level of achievement.

Although the curriculum for all subjects is set out in the same detail (in contrast with England where only core subjects are specified in detail) it is inevitable that the focus on assessment in the core subjects means that others receive less attention. There are reports that the quantity and quality of teaching in non-core subjects, particularly science (Watkins 2014; Estyn 2013), give some cause for concern.

## **Northern Ireland**

### ***Background***

Although a smaller country than Wales, Northern Ireland (population 1.8 million), like Scotland, has a long tradition as a separate education system. School education begins at the age of four and is compulsory to the age of 16.

The administration of the education system in Northern Ireland is complex, with 10 Statutory Bodies involved in the management and administration of the system. Among these are: The Department of Education (DE); the five Education and Library Boards and the Council for Catholic Maintained Schools (CCMS). The Council for the Curriculum, Examinations and Assessment (CCEA), established in 1994, is a non-departmental public body reporting to the Department of Education.

School education in Northern Ireland has six phases:

- Foundation (Years 1 and 2, ages 4 and 5)
- Key stage 1 (years 3 and 4, aged 6 and 7)
- Key stage 2 (years 5, 6, and 7, ages 8 – 11)
- Key Stage 3 (years 8, 9 and 10, ages 12 – 14)
- Key Stage 4 (years 11 and 12, ages 15 and 16)
- Sixth form (years 13 and 14, ages 17 and 18).

While most schools are grant-aided, there are 17 independent schools. There are 28 Irish-medium primary schools and one post-primary and one independent school teaching through the medium of Irish.

Although the 11 plus examination used to select pupils for secondary schools was officially ended in 2008, grammar schools continued to select by commissioning their own entrance examinations in English and mathematics. These tests, no longer provided by the CCEA, were purchased from tests developers and based on the curriculum for KS2. However, in 2013 the Department of Education issued guidance to schools and School Boards on transfer procedures, which they are legally required to take into account in drawing up their

procedures. This guidance explicitly states that schools do not use entry criteria related to academic ability.

### *Primary curriculum and assessment*

The Northern Ireland Curriculum for the primary school (Foundation and KS1 and 2) is set out in the same way for each of the areas of primary education. The areas of learning are

- Language and literacy
- Mathematics and numeracy
- The arts
- The world around us
- Personal development and mutual understanding
- Physical education
- Religious education

Specified outcomes include cross-curricular skills (communication, using mathematics and ICT) and thinking skills and capabilities (managing information, being creative, thinking, problem solving and decision making, working with others, and self-management).

The requirements for KS1 and 2 for each area of learning and for strands within each area are expressed as a list of what 'pupils should be enabled to know, do or understand.' In the case of science, history, geography and technology, which contribute to learning in 'The world around us', the statutory requirements are expressed in terms of 'what pupils should be enabled to explore'.

### *Purposes and methods of assessment*

The formative use of assessment not only receives general support but is particularly well developed and represented in the resources available to teachers. The Assessment for Learning Action Research Project, 2004-7, involved a number of primary schools and led to the development of *Assessment for Learning: A Practical Guide* (CCEA 2009). This substantial guide promotes a whole school approach to implementing assessment for learning, providing information and examples about the main formative assessment strategies of effective questioning, feedback, and scaffolding reflection including self-evaluation.

Schools are required to provide summative assessment data and report pupils' progress in the cross-curricular skills (communication, using mathematics and using ICT). At Key Stages 1 and 2, teachers assess and report with reference to the Levels of Progression for these skills, using the description of levels 1-5. (At KS3 a full range of levels 1-7 is used.) For Communication and Using Mathematics schools are required to report outcomes to parents and to CCEA at the end of each key stage. From 2013/14 revised moderation processes for Communication and Using Mathematics are in place to quality assure outcomes.

To assist teachers with their assessment of cross-curricular skills, a 'Task and Exemplification Library' is provided. For Communication and Using Mathematics the tasks are intended for use with Year 4 and Year 7 pupils for statutory assessment of these skills.

Teachers can choose appropriate tasks and find guidance about using tasks and levelling pupil work. The library also contains a variety of Using ICT tasks, which can be used with any year group to develop pupils' ICT skills as well as for statutory assessment.

For reporting to parents schools must use the Annual Pupil Profile which has been designed to promote consistency across Northern Ireland and is also used by Special Schools, Special Units attached to mainstream Primary Schools and Irish-Medium Schools and Units. By requiring reporting on all areas of the curriculum it avoids a narrow focus on reading, writing and mathematics.

There is also the Pupil Profile Summary which brings together for parents of Year 7 pupils all of the information gathered during the pupil's time in primary school.

## **New Zealand**

### ***Background***

New Zealand, with a population of about 4.5 million, is quite similar to Scotland in having a high proportion of small primary schools. School attendance (or equivalent) is compulsory from the age of 6 but most children begin at 5. The education system is organised in three phases: pre-school (up to the age of 5); primary (ages 5-13); secondary (ages 13-18). In some areas the top two years of primary are in intermediate (middle) schools. There is no selection within or at the end of primary education; secondary schools provide for the full range of student achievement. A separate sector of the system provides Maori-medium education that aims to promote the values of Maori teaching and learning philosophies and the use of the Maori language (Nusche 2012).

### ***Primary curriculum and assessment***

The New Zealand Curriculum, revised in 2007 and implemented in all schools in 2010, sets achievement objectives at eight overlapping levels covering Years 1 -13, but schools are not required to report on student performance in terms of levels. The curriculum identifies 'learning areas': English, the arts, health and physical education, learning language, mathematics and statistics, science, social science and technology.

National Standards (statements of expectations for student achievement and progress) are identified for the core subjects of mathematics, reading and writing. The standards set out, for the strands within these subjects, what pupils will have achieved: 'after one year at school'; 'after two years at school'; 'after three years at school'; 'by the end of year 4'; etc.

Teachers are expected to assess student performance against National Standards and report to parents regularly (at least twice a year) on their children's progress. Schools report on their pupils' performance in relation to the standards as part of their annual report to the Ministry of Education. This report is in terms of the number of pupils 'at', 'above', 'below' and 'well below' the National Standards for each year, but schools are not required to use these categories in reporting to parents. Rather it is expected that this should give detailed feedback, not numerical marks or any form of labelling. When assessing a pupil's

achievement and progress, the teacher needs to make an overall judgement about the student in relation to the whole standard.

### *Purposes and methods of assessment*

For years 1-10 all student assessment takes place within the school. There are no externally produced tests or required assessment tasks. Schools are free to decide on their approach to assessment and the tools they use. They can, however, make use of a bank of nationally validated assessment tools in making their own professional judgements.

There is a strong focus on the formative use of assessment in all documents for primary teachers. The guidance from the Ministry of Education presents assessment as integral to good teaching and learning, as less focused on summative judgements and more on decisions aimed at improving learning. The importance given to helping all pupils achieve is regarded as a major strength as is the role given to pupils in assessment:

New Zealand's assessment strategy...is characterised by an important devolution of assessment, starting with the students themselves. It emphasises the development of students' own capacity to regulate their own learning through self-assessment. (Nusche *et al*, 2012: 42).

Nusche *et al* (2012) also note that increased ability of pupils to talk about their learning makes it easier for teachers to find out about and support their pupils' learning. This approach is based on a firm belief in teachers' professionalism and underpinned by initial teacher education programmes which provide a basic preparation. Also, the existing culture of assessment in schools means that on graduation newly qualified teachers 'are rapidly involved in conversations with their colleagues about assessment through moderation activities and other professional activities associated with the school self-review and assessment of students' learning' (*op cit*, 43).

Nevertheless there is a warning that primary schools may overemphasise formative assessment to the detriment of useful summative assessment. The New Zealand inspectorate, the Education Review Office (ERO), in their 2007 report noted that while teachers collected 'accurate and valid' information in relation to English and mathematics, many did not do so in other curriculum areas (ERO 2007). This is perhaps not unexpected given that detailed standards exist only for the core subjects.

To help teachers with their assessment the Ministry of Education provides a guide to selecting appropriate assessment tools. The 'Assessment Tool Selector' lists the tools most frequently used in all areas of the curriculum, which include some tools created for the NEMP and NMSSA projects (see below). Also provided is the Observations Survey for Early Literacy, a set of tasks administered one-to-one designed to check a pupil's basic reading and writing concepts so that early intervention can be put in place if needed. For the core subjects the Assessment Tool Selector describes a range of more formal assessment tools for making judgement in relation to National Standards, whilst disclaiming any Ministry endorsement.



## *National monitoring*

The low stakes nature of individual student assessment at the primary level is preserved by the existence of a national monitoring programme. From 1995 to 2010 sample surveys were conducted by the National Education Monitoring Project (NEMP), funded by the Ministry of Education. In 2012 NEMP was replaced by the National Monitoring Study of Student Assessment (NMSSA) with similar aims and procedures but revised to take account of the change in the curriculum and in national and international survey models. Unlike other national surveys in England and Scotland, it uses a range of assessment modes used and covers all subjects not just those identified as 'core'.

Each year a sample of pupils at Year 4 and Year 8 take part and two learning areas are assessed (in 2012, science and English writing; in 2013, mathematics and health and physical education). Four modes of assessment are used: individual written responses, one-to-one interview, group work on performance tasks, and individual work on performance tasks. Trained teachers from other schools visit schools to administer the performance tasks and interviews. All one-to-one interviews and practical tasks are videoed to free administrators from making records at the time and are scored later by trained teachers. In NMSSA the whole national sample takes the individual written tasks, with the other three types of task administered to subsamples of the whole sample.

Whereas NEMP reported results item by item without any overall scores, NMSSA reports scores scales developed using a Rasch model, the Partial Credit Model (Masters, 1982). In the case of science, one of the first two subjects to be surveyed in 2012, the IRT software package WINSTEPS (Linacre, 2009) was used to develop the science scales. Year 4 and year 8 are on the same scale but different scales used for the two components of science: 'Knowledge and Communication of Science Ideas' and 'Nature of Science'. Differences among pupils by gender, age, ethnicity school type and school ranking are reported for these two science components.

## **Sweden**

### *Background*

Sweden, with a population of about 9 million, has a highly decentralised education system. Responsibility for organising and operating schools is vested in the 290 municipalities and independent providers. Attendance at school is compulsory for all children starting at the age of seven up to Year 9 (age 16). Since almost all children also attend non-compulsory primary school at the age of six, in practice children have ten years of education.

Municipalities have an obligation to provide nursery schools for children from one to five years of age whose parents work or study. Over 80 per cent of one to five year-olds spend part of their weekdays at nursery schools. The Swedish tradition of nursery school emphasises the importance of play in a child's development and learning and gender-awareness is an increasing feature.

Elementary school (Years 1–3) is followed by middle school (Years 4–6) and junior high school (Years 7–9). Almost all pupils then proceed to some form of upper secondary education programme lasting three years. Children between six and 13 are also offered out-of-school care before and after school hours. This can be at an after-school centre, a family day care home or an open after-school programme. Since 1997 Swedish schools have served free school lunches prepared at the school.

### *Primary curriculum and assessment*

The National Agency for Education (NAE) is the central administrative authority for the public school system, publicly organised preschooling, school-age childcare and adult education. It works to support municipalities and schools in achieving the curriculum goals laid down by the government. Within this framework the municipality or the board of an independent school may determine how its schools are to be run. Each school has the responsibility of working within the framework to achieve the goals but, due to the decentralised system in which municipalities interpret the national curriculum for their schools and school develop their own work plan, there is considerable variation among schools. An aim of the national curriculum introduced in 2011 was to provide more precise and detailed goals to help reduce this variation.

Pupils are given a part in developing specific goals for their school for each subject, based on the local goals and national curriculum. Even in the earliest grade levels, teachers discuss the goals and performance criteria with their pupils at the beginning of the year. Teachers are obliged to ensure that pupils and parents are well informed about the goals and receive regular feedback about progress through individual development plans (IDPs). (Swedish Ministry of Education and Research 2010).

The NAE is responsible for the national assessment system, and together with universities produces national tests and assessment guides for teachers. It is responsible for collecting data from schools in order to report on pupils, educational outcomes and costs. It also provides a national school leadership programme, recognising the importance of decisions made within the school.

### *Purposes and methods of assessment*

As in New Zealand, there is a strong focus on formative assessment and pupil participation in assessment. All assessment in the primary school is carried out by teachers. Pupils do not receive grades before Year 6 (until 2011 this was Year 8), but teachers report to parents at least twice a year. Individual development plans (IDPs) are an important feature of the implementation of formative assessment, specifying the individual learning goals so that progress towards them can be assessed and strategies designed to address any shortfall. They also serve to develop pupils' own assessment skills and are a basis for discussions with parents, held generally once a semester. The exact form of the IDP and the range of its content is decided within the school. In assessing pupils' progress teachers are expected to use a range of sources of evidence. There is also a 'test bank' for teachers to use if they wish.

The IDP and the results of national assessment (see below) can be used formatively, but on a long-term or medium-term timescale rather than the moment by moment or lesson by lesson, short-term timescale (Wiliam 2006). For the short-cycle interactions between teacher and pupils that have the most direct impact on achievement (Black and Harrison 2004; Looney 2011), formative assessment is an integrated part of the teaching and learning process. It will be incorporated into teachers' planning, questioning, dialogue with pupils, the form of feedback given to pupils and the encouragement of pupil self- and peer-assessment. Thus Nusche *et al* (2011) note that, although Sweden has a firm foundation for formative assessment, it can be much further developed in relation to short-term decisions.

For summative grading of pupils' overall performance from Year 6, there are learning goals for each subject which outline what is needed at different grade levels, A to F. Criteria are specified for grades A, C and E, with B and D applied by reference to the criteria for grades above and below for that year. The judgement is made on whether the majority of criteria for the grade above have been met. This approach means that pupils' grades may go up or down from one year to the next according to their achievement of the goals for each year.

### *National tests*

National tests are compulsory in Years 3, 6, and 9. In Years 3 and 6 the tests are in Swedish and mathematics and in Year 6 also in English. The tests are administered and marked by the pupils' own teachers and the results are used to inform judgements about overall grades. Although this provides teachers with information about the strengths and weaknesses of individual pupils, the timing of the testing, near to the end of the year, means that there is little opportunity for formative and diagnostic use of the results. However, this is just one function that the tests are expected to serve. On their multiple uses, Nusche *et al* (2011) comment:

The national tests currently serve many different functions including diagnostic, formative and summative assessment of individual students and they also produce the basic data for school self-evaluation, inspections and system-level evaluations. While the tests were originally designed to help teachers calibrate grades, they are increasingly used as national outcome measures. It is questionable whether in their current format the national tests can successfully fulfil all these expectations. (Nusche *et al* 2011:39)

There are several aspects of the national tests that raise concern about their usefulness in particular for national monitoring. Firstly, the marking by teachers introduces errors that make them unreliable. Secondly, since they are intended for diagnostic, among other, purposes they include performance items, which are generally of lower reliability and generalisability than standardised assessment (Linn *et al* 1991). Thirdly, there are simply not enough items to cover the subjects adequately, thus introducing a significant sampling error (see Wiliam 2001). Consequently the OECD report on evaluation and assessment in Sweden recommended that a national sample survey be considered for monitoring system-level outcomes and providing information about trends over time.

## France

### *Background*

In France (population about 65 million) the compulsory years of education are from six to 16 years. The school system comprises primary schools for children aged six to 11 years, middle/lower secondary schools for pupils aged 11 to 15/16 and upper secondary or Lycées for 15/16 to 18 year olds. Most children also attend kindergarten from the age of three or four. About 80 per cent attend state schools. However private schools are not equivalent to those in the UK since their teachers' salaries are paid by the state and they charge only low fees. Most are Catholic schools and teach religious education but also follow the national curriculum.

There is no selection or certification at the end of primary school but a contentious feature of French education is the high proportion (about 38 per cent) of pupils who repeat a year of schooling. This happens particularly in the final year of primary school if a pupil is considered not ready for middle school, often on account of a poor grasp of French.

### *Primary curriculum and assessment*

The Common Base of Knowledge and Skills (national curriculum), established in 2006, sets down what is to be taught and what must be learned in the primary and middle school years. For the primary years the curriculum is divided into three 'cycles' (the equivalent of key stages). These are: the Early Learning Cycle (kindergarten) the Basic Learning Cycle (final year of kindergarten and first two years of primary school), and the Consolidation Cycle (final four years of primary school).

In general terms the curriculum for the Early Learning Cycle identifies the major domains of activity to be covered over the three years which precede the start of compulsory schooling. It does not prescribe time to be spent on various activities but indicates that the experiences go beyond just play and include initial work in reading and number as well as art and creative activities. In the second cycle the priority objectives are learning to read and write in the French language, knowledge and understanding of numbers, writing numbers in figures (decimal numeration) and arithmetic using small quantities. Achievement in these domains is a general theme throughout all other activities. In the third cycle mastery of French and basic elements of mathematics continue to be the main objectives, with attention also given to expression in a modern foreign language and other subjects common to most primary curricula.

For the first cycle the curriculum establishes what pupils should be able to do at the end the kindergarten in respect of preparatory skills in reading and writing, social skills, learning about the world around, etc. For the second and third cycles the curriculum sets out in considerable detail the goals to be achieved in terms of what pupils should be able to do at the end of each cycle for each subject and strand within it. In addition, for French and mathematics, there are more precise objectives for each year. There is also specification of time to be spent on French and mathematics each week and on other subject annually. This annual, rather than day-to-day, allocation of time contrasts with the previous specification of

when and what to teach, leading to the reputation of French schools as all teaching the same topics at the same time. Thus the new framework provides flexibility allowing teachers and school teams to organize their teaching in cross-curricular topics and adjustable weekly or monthly blocks if they wish. Nevertheless the content of the curriculum remains quite tightly specified although pedagogical methods are left to teachers to decide.

### *Purposes and methods of assessment*

In the primary school on-going assessment by teachers has more of the nature of frequent summative assessment than is seen in New Zealand and countries of the UK. This is shown in the feedback to pupils, which assesses their work on the same criteria and is often in the form of marks. Raveaud (2004) reported that French teachers expected all pupils to tackle the same work rather than to be given easier work and risk being labelled a failure by their classmates. To teachers in England, facing pupils with tasks they are unable to do would be seen as damaging to their self-esteem, but this concern is overridden in France by a desire to give all pupils the same chance. This may well explain why repeating a year is so common in France and so rare in England. There is little written feedback of a formative nature, although in their oral remarks to pupils French teachers recognise effort and prior achievement.

Summative assessment in reading and writing in French and in mathematics is expressed as achievement in relation to the objectives set out for each year in the Common Base of Knowledge and Skills. Judgements are in terms of 'achieved' or 'not achieved'. For other subjects it is the responsibility of the school to decide the criteria, using help provided by the Ministry. A bank of test items covering most subjects has also been made available at a variety of levels for both primary and middle schools.

### *National tests and monitoring*

All pupils are tested on entering the third year of the primary school and the first year of the middle school in French, mathematics and science. The tests are provided by the Ministry of Education and are administered and marked by teachers. The purpose is diagnostic for teachers and parents, which is why they are conducted at the beginning of the year. The tests are based on the national curriculum and the outcomes are used to identify pupils' educational needs. Schools analyse their results using specific software provided. Because of the timing, the teachers cannot be held responsible for the results and there is evidence that teachers genuinely use them to inform their teaching (Bonnet 1997, quoting Th  lot 1993). Representative samples of the results from these compulsory diagnostic tests are collected and analysed centrally to provide a national picture of achievement and benchmarks for teachers. The test results influence practice by drawing attention to areas of weakness across all schools which can be addressed by teachers, as well as guiding plans for individual pupils.

However, this is not the only data on national standards available to the Ministry. There is also a national survey of samples of pupils at the end of primary and of middle school. As well as tests in all subjects, information is collected in these surveys about non-cognitive attainments, attitudes and values. Comparisons over time are made possible by including

some common items from year to year. Results for individual schools are not centrally reported; the survey findings are used centrally to report regional and national results. These results are widely distributed and used at regional level to identify those areas of the curriculum where schools may need help through professional development. At the national level resource allocation to regions takes into account the pupils' results in order to compensate for under-achievement which may have been caused by differential socio-economic factors (Bonnet 1997).

For school evaluation and accountability the emphasis is on self-evaluation against a set of national standard indicators, which include, but are not dominated by, pupil achievement data. Using the computer programs provided, schools are able to compare their profile with that of similar schools nationally or in the same region. Thus school evaluation does not rely solely on pupils' performance in tests or examinations.

### **Themes running through the examples**

Even across this limited selection of six countries outside England, there are some noteworthy themes that indicate the pros and cons of alternative assessment systems. A convenient structure is to consider themes relating to how the various systems provide for assessment for the main purposes of helping learning (formative assessment), reporting learning (summative assessment), contributing to the evaluation of teachers and school (accountability) and monitoring achievement at regional or national levels.

#### ***Formative assessment***

All of the systems reported here either implicitly or explicitly encourage the use of assessment to help learning. However, there is no statutory requirement for formative assessment, in recognition that it is a feature of teaching and the relationship between teacher and pupils that cannot be mandated. Of course it can be encouraged through non-statutory guidance and through the criteria for school self-evaluation. The specific help given to teachers in the countries considered is most detailed in Scotland, New Zealand and Sweden. In these systems there is an expectation that pupils will be given a role in the decisions about their work and so they will need to know the goals to work towards and the criteria by which the quality of their work is judged by the teacher and by themselves. In Scotland, pupil participation in assessment is explicit in the way that 'experiences and outcomes' in the Curriculum for Excellence are expressed. In Sweden the rule against giving pupils marks before year 6 means that feedback has to be via comments, which makes it more likely that the comments are formative, although this cannot be assured. Indeed a criticism of the formative assessment practice (Nusche *et al* 2011) was that it needs to be developed in relation to short-term decisions.

#### ***Summative assessment***

In relation to summative assessment for internal uses, for school records and reporting to parents and receiving teachers, there is considerable diversity in practice across the seven systems. A common element is that they all make use of assessment by teachers, but the way in which this is supported and reported varies considerably. England is the only

country using externally marked tests in primary education. Wales, Sweden and France require tests in aspects of language and mathematics, but these are marked by teachers and have a role in informing teachers' assessment. In Scotland, Northern Ireland, France, Sweden and New Zealand, teachers are provided with banks of activities and tools which they can choose to use in making their summative judgements.

The form of reporting varies markedly. In England the intention is to report test results for pupils aged seven and 11 as scaled scores, introducing a normative element, in sharp contrast to the essentially criterion-referenced assessment in other countries. This is compounded by providing information to parents as to how their child's score compares with the average for the school, local area and nationally. These scores are for aspects of English and mathematics. For other subjects, as in France, the form of reporting is left to schools and teachers to decide.

There are significant differences among countries concerning the use of 'levels' in reporting pupils' achievement. In New Zealand, even though objectives (standards) are set out at eight overlapping levels covering Years 1-13, levels are not used in reporting on individual pupils. In England the use of progressive levels has been abandoned in the arrangements being introduced from 2014, for reasons mentioned earlier. However, levels of this kind have been retained for reporting in Wales and Northern Ireland. In Scotland, 'levels' refer to the stages of schooling at which the 'experiences and outcomes' are identified. These statements express the expected outcomes at the end of each level, or stage, and pupils' performance is reported in terms of the extent of achievement of a level ('developing', 'consolidated' or 'secure'). An alternative to reporting in terms of progressive levels which incorporate a description of performance, is the use of standards which specify outcomes for each year, as in New Zealand and France. Teachers judge whether a pupil has 'achieved', or 'not achieved' (in France), or is performing 'at', 'above' or 'below' (in New Zealand) the standard for the year. In Sweden a similar approach is used after Year 6 but with performance against each year's goals being graded A to F. In England, levels have been replaced by performance descriptors which are used by teachers to report performance in the core subjects or mathematics, reading, writing and science. At KS1 teachers use judgements about performance across a number of elements to decide which of four standards best matches a pupil's overall attainment: mastery standard; national standard; working towards national standard; below national standard (see page 11).

Parents receive an annual report on all areas of the curriculum in most countries but the concern expressed by the New Zealand inspectorate about the dependability of summative assessment in subjects other than language and mathematics may well apply to other countries. A different degree of specification of the curriculum content for core and non-core subjects not only provides less help for teachers in assessing achievement in the non-core subjects but signals that these are less valued. The requirement for moderation of teachers' assessment to be used only for the skills of communication and using mathematics, as in Northern Ireland, might convey the same message.

### *Accountability and school evaluation*

The use of pupil assessment data as a measure of the quality of teaching is a matter of

contention. It can be argued that schools and teachers should be judged only on the actions and outcomes for which they are accountable. In the context of pupils' learning, teachers can be held accountable for what they do in the classroom, what learning opportunities they provide and the help they give to pupils, and so on. They are not necessarily responsible for whether externally prescribed learning outcomes are achieved, since this depends on other factors over which the teacher does not have control, such as the pupils' prior learning and the many out of school influences and conditions that affect their learning. Thus teachers and schools ought to be held to account for the programme of learning opportunities that is provided and the evidence of relevant learning, but not judged solely on the level of outcomes reached by their pupils, particularly when these tests are restricted to literacy and numeracy.

In many countries there is opportunity for the achievement of pupils to be reported in the context of the many factors that affect this achievement. In Scotland, Wales and Northern Ireland for example, external school evaluation is complemented by internal school self-evaluation using guidelines provided by the inspectorate (e.g. HMIe 2007; Estyn 2004). This reflects an overall aim of the systems, found also in New Zealand, for the assessment of pupils and evaluation of schools to provide those in schools with tools to improve their practice rather than to be used by others to control teachers and schools.

The practice noted in France and England of providing schools with software that enables them to compare the performance of their pupils with those in schools of similar background is becoming more widespread. However, this encourages multiple use of data, a practice which was heavily criticised in England (Mansell *et al* 2009) and in the OECD report on assessment in Sweden. Not only is it questionable that the data can be equally valid for formative and summative purposes but using pupil performance data for school accountability makes no allowance for differences in pupil intake. The planned new procedures for school accountability in England attempt to recognise the different starting points of pupils entering school. Progress from the starting point of the reception baseline test to the end of primary school will be one measure used in evaluating a school. However, there is an alternative measure based only on the level of attainment in the Year 6 tests. This means that there is a choice for schools, to be judged on progress or on raw scores only, which Mansell (2014) suggests leaves opportunity for the manipulation of the accountability measure and is 'a recipe for confusion'.

### ***National monitoring***

In France, Scotland and New Zealand externally set and marked tests are used for the purpose of monitoring regional and national standards, but in all three cases these are quite separate from tests used for school and teacher accountability. These monitoring surveys use a matrix sampling design and involve a relatively small sample of pupils on each occasion. They are part of an ongoing programme designed to show not only what pupils can do at any one time but also to monitor changes across the years. A large number of items can be used in a survey, with any one pupil taking only a few of those used in any one survey. Thus together the survey items can provide a good sample of the curriculum content although only in New Zealand does the range extend beyond the core subjects. In England, sample tests are carried out in science only.



National surveys of the kind used in Scotland, New Zealand and France, divide the total number of items into sub-tests each given to a random sample of the population. This is the design used in international surveys of TIMSS and PISA and in national surveys such as the NAEP in the USA and in the APU when it operated in England, Wales and Northern Ireland in the 1980s (Foxman *et al* 1991). The purpose of these surveys is to identify performance at national or regional levels, which is derived from the combination of results from the sub-tests. The results for individual pupils in the sample, or even for a whole school, have little value in terms of the overall goals being assessed, but when combined across the whole national sample, the results provide a much better picture of national performance than would be obtained if every pupil took the same test items. When different items are given to different student samples, there is no need to cram into a short test a number of short items and more time can be given for pupils to engage with a particular context. Results at the class and school level are of value only when combined with others to report at regional or national levels and so cannot be used for school evaluation. These surveys are therefore described as having low stakes.

In England the APU was terminated in favour of using the results of national tests, which were used to set targets and rapidly acquired high stakes. Thus now in England, Wales, Northern Ireland as in Sweden, national performance in English and mathematics is computed from the aggregated score of all pupils on the same test items. The low reliability and validity of these tests and the high stakes uses of the scores at school level means that they provides a poor measure of the quality of learning and teaching.

Wide reporting of results of national sample surveys enables the information to be used formatively at the system level, providing feedback that can be used to identify aspects of the curriculum that may need attention. The value of this information to schools is in focusing attention on their own practice and the performance of pupils in the areas identified as weaknesses nationally. This use of the results encourages participation in the surveys. In this way national data are collected without adding high stakes to the assessment of pupils.

High stakes use of tests is also avoided by ensuring that the evaluation of teachers and schools for accountability is based on a range of indicators relating to the context, environment, curriculum provision and resources as well as pupil performance. In Scotland, France, and Wales, such varied indicators are provided for school evaluation and school self-evaluation (Estyn 2004; HMIe 2007).

## 5 – DISCUSSION: THE SYSTEM IN ENGLAND

Even if we do not wish to go so far as to claim that what is assessed determines what is taught, it cannot be denied, as stated at the start of this report, that assessment does have a large impact on pupils' education experiences. Indeed the OECD (2013) reports an increasing focus on assessment of pupils and evaluation of teachers, schools and education systems. Its review of evaluation and assessment in 28 countries identified four main factors driving this trend:

- An increased demand for effectiveness, equity and quality in education to meet economic and social challenges.
- A trend in education towards greater school autonomy, which is fuelling a need to monitor how schools are doing.
- Improvements in information technology, which allow for the development of both large-scale and individualised student assessment and facilitate the sharing and management of data.
- Greater reliance on evaluation results for evidence-based decision making. (OECD 2013, Executive Summary: 1)

These trends can be identified in some of the assessment systems outlined in section 4. In addition the OECD review identified several policy challenges. A selection from these, relating to pupils and of most relevance here, are:

- Aligning assessment with educational goals and learning objectives set out in the curriculum.
- Focusing on improvement of classroom practices and building on teacher professionalism.
- Designing the accountability use of assessment in ways which minimise undesirable effects.
- Placing the student at the centre, fostering engagement in learning through using formative assessment strategies.

These points offer criteria against which the practice in England, and potential alternative practice, can be judged.

In relation to the first point it is assumed that 'the learning objectives set out in the curriculum' is intended to refer to the whole National Curriculum in the case of England; that is, core and non-core subjects. The amount of detail in which both the programme of study and the assessment standards are expressed for the core subjects is in sharp contrast to the very general treatment of these matters for the non-core subjects. However, even for the core subjects it is not easy to judge the alignment of the assessment with the intended outcomes, because the national curriculum states what pupils are 'expected to know, apply, and understand' for each year, while the statements relating to the national standards apply only at the end of each key stage. An exact match is not, then, expected but there should be a broad correspondence between the curriculum and the elements used by teachers in judging the standard reached by pupils. The amount of detail given in the 50 or so elements proposed for judging each standard tends to obscure the overall picture of what it means to attain a particular standard. Further, since there are no performance descriptors for any subjects other than mathematics, reading, writing and science, the system falls short of providing measures relating to the full range of pupils' learning.

In challenging policy to ensure that the impact of evaluation and assessment improves classroom practice the OECD report states that 'all types of evaluation and assessment should have educational value and should have practical benefits for those who participate in them' (OECD 2013 Executive Summary: 2). The principle that 'Assessment of any kind

should ultimately improve learning' was also emphasised by Gardner *et al* (2008) after studying how changes in assessment practice in primary classroom can be brought about most effectively. The impact on learning of implementing formative assessment is obvious, for unless this is positive the assessment cannot be described as formative (Black *et al* 2002). It is much less clear in the case of assessment for monitoring at school or national level, where results are aggregated and there are no direct consequences for the pupils involved. However, even in this case the result should ultimately improve learning of future pupils by being used to inform changes in curriculum teacher education or other aspects of provision.

In between the close use of assessment to help individual learners in the classroom and the very distant use for monitoring and for policy decisions are the uses of summative assessment for recording and reporting pupils' achievements. For reports to be meaningful to parents they need to summarise what children can do and understand. The arguments against using levels put forward by the government's expert group, included concern that a level does not convey this detailed information, although it can indicate progress from year to year. But now, instead of levels, results for aspects of language and mathematics tests will be reported as points on a scaled score, set with the average at 100. These scores indicate only where a child is in relation to others and nothing about where they are in progression in learning. Hopefully this will be complemented by more substantial information about subjects not included in tests, using reporting procedures proposed by teachers themselves and providing a profile as used in several of the countries included in section 4.

The accountability measures to be introduced leave open a number of opportunities for 'undesirable effects'. There are two ways in which schools can reach the 'floor standard' and thus avoid the penalties that falling below this will bring. One is the progress standard, measured by the gain in pupils' scores between the 'reception baseline' and the end of key stage two tests. The other is the 'attainment standard', reached if 85 per cent of pupils achieve the expected standard. The high stakes for schools attached to these measures opens up the possibility of dubious practices, such as depressing pupils' baseline results in order to inflate the measure of progress. Even without this, the validity of the results from testing such young children within the first few weeks of starting school is highly questionable, quite apart from the effects on the pupils. The alternative accountability measure, related only to the KS2 test results, will be seen as attractive to schools with an intake of children who are performing well and are likely to meet the required standard without extra effort.

For either measure of the floor standard the pressure on performance in the key stage two tests is greater than in the previous system, particularly as the standard to reach has been raised: 85 per cent of pupils will have to reach the level previously reached by 65 per cent (DfE 2014: 21). This standard has to be reached in all of reading, writing and mathematics, whereas previously this applied to any one or more of these three. The reason given for this requirement being that it ensures that children 'succeed across the curriculum'. Whilst it is true, of course, that these subjects are fundamental to all learning in the primary school it is hardly justified to infer that this alone is sufficient.

The OECD report's challenge to 'place the students at the centre' is consistent with the insistence that assessment should help learning. Formative assessment, as indicated in section 2 and Figure 1, has the learner at the centre, fully engaged in the decisions about the

steps needed to achieve the lesson goals. Pupils in any case are the ones who have to take actions that lead to learning, but whether or not they take responsibility for these actions depends on their understanding of why these actions are necessary and the ability to recognise when they have been accomplished. This was the reason for the system in Sweden giving pupils a role in defining goals as well as in participating in assessment of progress towards them.

Giving pupils opportunity to talk about their learning has several benefits, including encouraging the habit of reflection on the process of learning, promoting higher order thinking and meta-cognition. Also, as noted in the review of practice in New Zealand, giving opportunity to talk about their learning makes it easier for teachers to find out about and support their development. Insisting on 'no marks' assures that the discussion is meaningful in terms of the substance of what is being learned. Moreover, having a role in their assessment means that pupils view it as a process in which they have a part, not something that is 'done to' them.

### **Alternative approaches**

The new system for assessment and accountability for primary schools in England still suffers from over dependence on testing and the use of end of key stage two tests for too many purposes. End of key stage tests results are used, as a minimum, for reporting to parents and to receiving secondary schools, for accountability, judging whether schools are above the 'floor standard' and for monitoring national performance. They cannot be well matched to all these purposes and any bias due to any one use will influence the validity for other uses. There is therefore need to consider alternative approaches.

For end of key stage reporting, whilst the greater use of teachers' assessment is welcomed, the use of the proposed performance descriptors by teachers to make their judgements has two main limitations. First, the approach only concerns reporting in the core subjects. Second, there is a risk that the highly detailed and long lists of statements defining each standard will be treated as a check-list, with the meaning of what has been learned lost in the detail. Judging the standard that has been reached requires multiple decisions such as whether a pupil uses 'a dictionary or thesaurus ... to check work meaning and appropriateness' (writing KS2) or 'describes the main changes as seeds and bulbs grow into mature plants' (science KS1). A more holistic approach would provide a more valid assessment of pupils' attainment in writing or understanding of organisms.

Rather than working through long lists of statements teachers ought to be able to use criteria that they can keep in mind and use in their everyday contact with pupils, so that relevant information is already available when it comes to decisions about attainment at the end of key stages. Various ways of moderating teachers' judgements can be applied to ensure confidence in the results so that they can be used for internal school monitoring. At the primary level the most appropriate form is group moderation involving teachers meeting together to discuss examples of pupils' work. These meetings enable teachers to share their interpretations of the performance descriptors as well as to discuss their judgements of specific sets of work. Experience of group moderation suggests that it has benefits beyond the quality of assessment results. It has a well-established professional development

function since discussing the inferences that can be drawn from studying pupils' work provides teachers with insight into the nature of the assessment process which benefits not only their summative assessment but also their formative use of assessment.

Information for regular reporting to parents and keeping track of pupils' progress as they move through the school should cover the full range of learning outcomes, not just the core subjects. The performance descriptors cannot be used to provide this information. Instead, evidence relating to each pupil's performance in relation to the school's programme of study for each subject can be accumulated over a term or year. Among the various approaches for doing this the most useful are those that enable pupils to have a role in the process by taking part in selection of items in their workbook or kept in a folder or portfolio of best work. Setting aside time for pupils to review their work and make the selection provides opportunities for self-assessment. The collection of work – which can take a variety of forms not confined to written work and may eventually be stored electronically – is itself a record of what the pupil can do and understand and a basis for communication with parents.

Black *et al* (2013) report work with teachers aimed at developing teachers' assessment so that it can meet the requirements of reliability and validity. Although the work was carried out with secondary teachers of English and mathematics, the approaches found effective are equally relevant in the primary school. For instance, asking teachers to discuss 'what does it mean to be good at (a particular aspect of the subject)' and then consider how to provide relevant information identified as needed helped teachers to reflect upon and refocus their assessment. The collection of assessment tasks in a portfolio was found to be the most effective basis for dependable summative reporting.

For effective and equitable accountability, information is also needed about all areas of learning, as well as about learning opportunities provided and other facts and features that influence pupils' learning. Some of this will be about pupils' learning and based on moderated teachers' assessment, which gives a better picture of the school's performance than tests of language and mathematics. School self- evaluation, moderated by a balance of regional/area and national inspection services (Green *et al* 2007), will mean that the process is formative for the school.

Any system must make provision for national monitoring of pupil performance, which is undeniably an essential measure of the effectiveness of an education system. Using teachers' assessment is not suitable since the evidence base varies from pupil to pupil, so some form of test has to be used. For national monitoring, a far greater sample of performance in a domain is needed than is provided by collecting the results of individual pupils who have all taken the same test. As noted earlier, a small sample of the pupil population, between them answering a wide range of items, is all that is needed to provide a good estimate of pupils' performance in a domain and to identify where strengths and weaknesses lie to inform policy and practice. Although monitoring is restricted to aspects of language and mathematics in Scotland and France, and the only sample testing in England is in science, the practice in New Zealand shows that this is not a necessary restriction.

Finally, an effective assessment system is an open one, where all involved know what evidence is used and how it informs judgements. Much of the emotion aroused by

assessment is a result of fear or suspicion of the unknown. To take this away we need to be completely open about the need for and purpose of assessment and why it is carried out in particular ways. Even the youngest pupils can be given some explanation of what evidence they and their teachers can use to judge the progress they are making. This helps pupils to take part in assessing their own work but it is equally important for summative assessment so that there are no surprises (for pupils or parents) in the reports of what has been achieved at a particular time.

## 6 – IMPLICATIONS FOR ASSESSMENT POLICY AND PRACTICE

There are significant implications for assessment policy and practice in England to be drawn from this discussion. Before listing the main ones it is important to make explicit the basis of value judgements that inevitably underpin all such statements and are often expressed as guiding principles. The principles endorsed here are the following, drawn from Gardner *et al* (2010):

- Assessment of any kind should ultimately improve learning.
- Assessment methods should enable progress in all important learning goals to be facilitated and reported.
- Assessment procedures should include explicit processes to ensure that information is valid and is as reliable as necessary for its purpose.
- Assessment should promote public understanding of learning goals relevant to pupils' current and future lives.
- Assessment of learning outcomes should be treated as approximations, subject to unavoidable errors.
- Assessment should be part of a process of teaching that enables pupils to understand the aims of their learning and how the quality of their work will be judged.
- Assessment methods should promote the active engagement of pupils in their learning and its assessment.
- Assessment should enable and motivate pupils to show what they can do.
- Assessment should combine information of different kinds, including pupils' self-assessments, to inform decisions about pupils' learning and achievements.

These principles are reflected in the implications drawn for the assessment practice of those working in English education as classroom teachers, school leaders, policy-makers and teacher educators.

### *Classroom teachers*

- Teachers should use a range of assessment methods that enable the full range of goals of learning and progression towards them to be addressed and that do not restrict the breadth of the curriculum.
- Teachers use evidence from their on-going assessment to:
  - help pupils' learning;
  - summarize learning in terms of reporting criteria;

- reflect upon and improve their teaching.
- Teacher should use assessment to advance pupils' learning by:
  - using on-going evidence of learning to adapt the pace, challenge and content of activities;
  - giving feedback to pupils about how to improve or move forward;
  - providing time for pupils to reflect on and assess their own work.
- Teachers should base their summative judgements of pupils' learning outcomes on evidence drawn from a range of sources which might include tests as well as regular class activities.

### *School leaders*

- There should be a school policy for assessment that:
  - requires teachers to conduct effective formative and summative;
  - emphasises formative assessment with summative judgements made only when necessary to check and record progress;
  - is regularly discussed among teachers and shared with parents.
- Assessment should be conducted through methods consistent with pedagogy that promotes understanding and the development of thinking and learning skills.
- Teachers should be encouraged to specify in their lesson planning how to gather evidence of pupils' learning so that assessment is aligned to the curriculum.
- Time should be made available for teachers to meet and discuss assessment and on occasion to observe each other's practice.

### *Policy-makers*

- The current use of externally produced and marked tests should be replaced by moderated and well-supported assessment by teachers, which can provide information of greater dependability across all subject areas.
- Assessment criteria, linked to programmes of study, should be provided for all subject areas and sub-sections within them so that assessment can be used to help learning and provide a complete record of achievement across the full curriculum.
- Moderation to improve the reliability of teachers' summative assessment should be required, and action taken to overcome distrust through openness about procedures and wider recognition that any assessment is subject to unavoidable errors and results should be treated as approximations.
- Monitoring of national standards should be based on sample surveys using a large number of items within a rolling programme which extends beyond the core subjects.

### *Teacher educators*

- Initial teacher education and professional development should ensure that teachers have the skills to use assessment to support learning and to make reliable summative judgements for reporting pupils' attainment.
- The solid support in the academic literature and research for formative assessment should be reflected in the resources, including time, provided in teacher education

courses for developing practices such as questioning, providing feedback, sharing goals with pupils and encouraging pupil self- and peer-assessment.

- Group moderation of teachers' judgements of pupils' work should be recognised as a valuable form of professional development.



## REFERENCES

- Alexander, R. J. *et al* (2010) *Children, their World, their Education: final report and recommendations of the Cambridge Primary Review*. London: Routledge.
- Assessment Reform Group (ARG) (2002) *Assessment for Learning: 10 principles*. Online. Available: [www.assessment-reform-group.org](http://www.assessment-reform-group.org), and from the CPA office of the Institute of Education, University of London.
- Black, P., Harrison, C., Hodgen, J., Marshall, B. and Serret, N. (2013) *Inside the Black Box of Assessment*. London: GL Assessment.
- Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (2002) *Working Inside the Black Box*. London: GL Assessment.
- Black, P. and Harrison, C. (2004) *Science inside the Black Box*. London: GL Assessment.
- Black, P. and Wiliam, D. (1998a) *Inside the Black Box*. London: GL Assessment.
- Black, P and Wiliam, D. (1998b) 'Assessment and classroom learning'. *Assessment in Education* 5(1): 1-74
- Bonnet, G. (1997) 'Country profile from France'. *Assessment in Education* 4(2): 295-306
- CCEA (2009) *Assessment for Learning: A Practical Guide*  
[http://www.nicurriculum.org.uk/docs/assessment\\_for\\_learning/AfL\\_A%20Practical%20Guide.pdf](http://www.nicurriculum.org.uk/docs/assessment_for_learning/AfL_A%20Practical%20Guide.pdf)
- DfE (Department for Education) (2014) *Reforming Assessment and Accountability for Primary Schools*. London: DfE.
- DfE (Department for Education) (2012) *Statutory Framework for the Early Years Foundation Stage*. London: DfE.
- DfE (Department for Education) (2011) *The Framework for the National Curriculum. A Report by the Expert Panel for the National Curriculum Review*. London: DfE.
- Elton, L. (2004/5) 'Goodhart's Law and performance indicators in higher education'. *Evaluation and Research in Education*, 18(1 & 2): 120-128
- ERO (Education Review Office) (2007) *The Collection and Use of Assessment Information in Schools*. Education Review Office, Wellington.
- Estyn (2004) *Guidance on the Inspection of Primary and Nursery Schools*. Cardiff: Estyn.  
 Estyn (2013) *Science in Key Stages 2 and 3*. [www.estyn.gov.uk](http://www.estyn.gov.uk).
- Foxman, D., Hutchinson, D. and Bloomfield, B. (1991) *The APU Experience, 1977-1990*. London: Schools Examination and Assessment Council.

Gardner, J., Harlen, W., Hayward, L. and Stobart, G. (2008) *Changing Assessment Practice*.  
<http://www.aria.qub.ac.uk/>

Green, S., Bell, J., Oates, T. and Bramley, T. (2007) 'Alternative approaches to national assessment', unpublished paper.

Harlen, W. (2013) *Assessment and Inquiry-Based Science Education: Issues of Policy and Practice*. IAP. <http://www.lulu.com/content/paperback-book/assessment-inquiry-based-science-education-issues-in-policy-and-practice/13672365>

Harlen, W. (2012) 'On the relationship between assessment for formative and summative purposes', in J. Gardner (ed.) *Assessment and Learning*. London: Sage, 87-102.

Harlen, W. (2006) *Teaching, Learning and Assessing Science 5 – 12*, 4<sup>th</sup> edition. London: Sage.

Harlen, W. (2010) 'The quality of learning: assessment alternatives for primary education', in R.J.Alexander *et al* (ed) *The Cambridge Primary Review Research Surveys*, London: Routledge, 484-520.

HMIe (2007) *How Good is Our School? The journey to excellence*. Edinburgh: HMIe. Online (Available: <http://www.hmie.gov.uk/documents/publication/hgiosjte.pdf>).

Linacre, J. M. (2009) *WINSTEPS Rasch measurement computer program*. Chicago: Winsteps.com.

Linn, R.L Baker, E.L., and Dunbar, S.B (1991) 'Complex, performance-based assessment: expectations and validation criteria'. *Educational Researcher*, 20: 5-21.

Looney, J. (2011) 'Integrating formative and summative assessment: progress toward a seamless system?' *OECD Education Working Paper 58*, OECD: Paris. Available from [www.oecd.org/edu/workingpapers](http://www.oecd.org/edu/workingpapers)

Mansell, W. (2014) 'Primary and accountability proposals: what we've learned', blog on the NAHT website 31/3/14. See <http://www.naht.org.uk/welcome/news-and-media/blogs/warwick-mansell/primary-and-accountability-proposals-what-wevelearned/>

Mansell, W., James, M. and the ARG (2009) *Assessment in Schools. Fit for purpose? A Commentary by the Teaching and Learning Research Programme*. London: Economic and Social Research Council Teaching and Learning Programme.

Masters, G.N. (1982) 'A Rasch model for partial credit scoring'. *Psychometrika*, 47: 149-174

Messick, S. (1989) 'Validity', in R.L. Linn (ed) *Educational Measurement*, 3<sup>rd</sup> Edition. London: Collier Macmillan, 12-103.

Newton, P.E. (2012) 'Validity, purpose and the recycling of results from educational Assessment', in J. Gardner (ed) *Assessment and Learning*, 2<sup>nd</sup> edition. London: Sage, 264-276.

Newton, P.E. (2007) 'Clarifying the purposes of educational assessment'. *Assessment in Education*, 14(2), 149-170.

Nusche, D., Laveault, D., MacBeath, J. & Santiago, P. (2012) *OECD Reviews of Evaluation and Assessment in Education: New Zealand 2011*. Paris: OECD Publishing.

Nusche, D., Halasz, G., Looney, J., Santiago, P. and Shewbridge, C. (2011). *Reviews of Evaluation and Assessment in Education: Sweden*. Paris: OECD Publishing.

OECD (Organisation for Economic Co-operation and Development) (2013) *Synergies for Better Learning: An International Perspective on Evaluation and Assessment*, OECD Reviews of Evaluation and Assessment in Education. Paris: OECD Publishing.

Pine, J., Aschbacher, P., Rother, E., Jones, M., McPhee, C., Martin, C., Phelps, S., Kyle, T. and Foley, B. (2006) 'Fifth graders' science inquiry abilities: a comparative study of students in hands-on and textbook curricula'. *Journal of Research in Science Teaching* 43(5): 467-484.

Qualter, A. (1985) *APU Science Practical Investigations: a study of some problems of measuring practical scientific performance in children*. (Unpublished PhD thesis) University of London.

Raveaud, M. (2004) 'Assessment in French and English infant schools: assessing the work, the child or the culture?' *Assessment in Education* 11(2): 193-212.

Scottish Government (2011) *Curriculum for Excellence Building the Curriculum 5: A Framework for Assessment*. Edinburgh: Scottish Government.

SEED (2004) *Assessment, Testing and Reporting 3-14: our response*. Edinburgh: SEED.

STA (Standards and Assessment Agency) (2014) *Reception baseline: criteria for potential assessments* London: STA.

Stobart, G. (2008) *Testing Times. The Uses and Abuses of Assessment*. London: Routledge.

Swedish Ministry of Education (2010) *Country Background Report for Sweden*, prepared for the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes. [www.oecd.org/dataoecd/22/44/45957739.pdf](http://www.oecd.org/dataoecd/22/44/45957739.pdf)

Thélot, C. (1993) *L'Évaluation du Système Éducatif*. Paris: Nathan.

Torrance, H. (2007) 'Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning'. *Assessment in Education*, 14(3): 281-294.

WAG (Welsh Assembly Government) (2008) *Foundation Phase Framework for Children's Learning for 3 to 7 year olds in Wales*.

[http://wales.gov.uk/topics/educationandskills/earlyyearshome/foundation\\_phase/foundation-phasepractitioners/frameworkchildlearning/?lang=en](http://wales.gov.uk/topics/educationandskills/earlyyearshome/foundation_phase/foundation-phasepractitioners/frameworkchildlearning/?lang=en)

WAG (Welsh Assembly Government) (2010) *Making the Most of Assessment 7-14*.  
<http://wales.gov.uk/docs/dcells/publications/100511assessment714en.pdf>

Watkins, R. (2014). 'Curriculum change and raising standards: the Welsh perspective'.  
*Primary Science*, 133: 20-21.

William, D. (2006) 'Formative assessment: getting the focus right', *Educational Assessment*, 11: 283-289.

William, D. (2001) 'Reliability, validity and all that jazz', *Education 3-13* 29(3): 17-21.

William, D. (1993) 'Reconceptualising validity, dependability and reliability for National Curriculum Assessment'. Paper given at the British Educational Research Association conference, September 1993.

Wyse, D., McCreery, E. and Torrance, H. (2010) 'The Trajectory and Impact of National Reform: curriculum and assessment in English primary schools', in R.J. Alexander *et al* (ed) *The Cambridge Primary Review Research Surveys*, London: Routledge: 792-817.





...children, their world, their education

The Cambridge Primary Review Trust is the successor to the Cambridge Primary Review (2006-12). It aims to extend and build upon the Review to advance the cause of high quality primary education for all. It is supported by Pearson Education, based at the University of York, and chaired by Professor Robin Alexander.

#### **FURTHER INFORMATION**

[www.cprtrust.org.uk](http://www.cprtrust.org.uk)

General enquiries: [administrator@cprtrust.org.uk](mailto:administrator@cprtrust.org.uk)

Published by the Cambridge Primary Review Trust,  
University of York,  
York, YO10 5DD, UK

ISBN 978-0-9931032-0-9

Copyright © 2014 Cambridge Primary Review Trust