Nuffield Foundation

Developing policy, principles and practice in primary school science assessment

Report from a working group led by Professor Wynne Harlen
July 2012

Nuffield Foundation
28 Bedford Square
London WC1B 3JS
Telephone: +44 (0)20 7631 0566
www.nuffieldfoundation.org

The Nuffield Foundation is an endowed charitable trust that aims to improve social well-being in the widest sense. It funds research and innovation in education and social policy and also works to build capacity in education, science and social science research.

Contents

Foreword

The origins of this report lie in the 2012 annual conference of the Association for Science Education (ASE). The Nuffield Foundation – in collaboration with the Science Learning Centre Network, the Royal Society and the Wellcome Trust – organised a panel discussion addressing the question: 'Is it possible to design assessment for the science curriculum that is fit for purpose?'

For those with an interest in primary science, it was clear there was scope for further discussion, not least in the context of the review of the national curriculum. In response, Professor Wynne Harlen convened a working group that brought together specialists in primary science education and assessment. Members of the group spent time considering an appropriate framework for primary science assessment, culminating in a two-day seminar held at the Nuffield Foundation in June.

The group's main concern was the impact on primary science of the removal of national testing at the end of Key Stage 2 (age 11) in 2010. This left a vacuum in terms of the *type*, *purpose* and *use* of assessment at this level. Removal of formal testing has had implications for the perceived status of science compared with subjects that are still tested at this stage (English and mathematics). On the other hand, it has also had a positive effect in some ways, such as greater curriculum flexibility for teachers and less 'teaching to the test'.

This report sets out the group's proposed framework for the assessment of primary science. Based on an agreed statement of principles, the group has outlined how pupils' attainment should be collected, recorded, communicated and used. The framework emphasises and clarifies the role of assessment in the learning process, as well as highlighting the need for a range of methods.

The Nuffield Foundation is very grateful to Wynne and her colleagues for the work they have undertaken in the production of this excellent report. We hope this is the start of a discussion leading to a more considered and effective approach to science assessment in the primary curriculum.

Josh Hillman
Director of Education

The working group

Professor Wynne Harlen OBE (Chair)
Past president of the Association for Science Education (ASE); member of SCORE (Science Community Representing Education) Primary Committee; Primary School Quality Mark (PSQM) advisory group; and Wellcome Primary Science Specialist (PSS) programme external advisory group

Professor Derek Bell
Director, Campanula Consulting; College of Teachers; former CEO of ASE; former Head of Education at the Wellcome Trust; member of advisory board for PSQM

Marianne Cutler
ASE Director of Curriculum Innovation; member of SCORE

Anne Goldsworthy
Consultant; former Chair ASE Primary Committee; member SCORE Primary Committee

Dr Angela Hall
Director, Science and Mathematics Education, Nuffield Foundation

Dr Christine Harrison
Co-Director of Assessment for Learning and Senior Lecturer in Science Education, King's College London; former member of ASE Council

Sally Howard
Senior Teaching Fellow, Primary Science, University of Warwick; member of ASE Council and Assembly; ASE Trustee; Teach First Primary Programme lead West Midlands

Brenda Keogh
Millgate House Education; member of ASE Primary Committee, SCORE Primary Committee, and PSQM advisory group

Liz Lawrence
Barking and Dagenham Local Authority Adviser; ASE Chair-elect

Stuart Naylor
Millgate House Education; member of ASE Council

Professor Michael Reiss
Pro-Director: Research and Development and Professor of Science Education, Institute of Education, University of London

Professor Dudley Shallcross
University of Bristol; Director of the AstraZeneca Science Teaching Trust

Jane Turner
Leader, PSQM award programme; Associate Director, Science Learning Centre East of England; member of ASE Primary Committee and ASE Assembly; Lead Science Curriculum Expert DfE Standards and Testing Agency

Executive Summary

This report sets out a proposed framework for the assessment of science in primary schools (Key Stages 1 and 2).[1] The framework is the outcome of a seminar held at the Nuffield Foundation on 7 and 8 June 2012. Participants included representatives of the Association for Science Education (ASE), SCORE (Science Community Representing Education), the Science Learning Centre East of England, the AstraZeneca Science Teaching Trust and the Primary Science Quality Mark project.

The framework describes how evidence of pupils' attainment should be collected, recorded, communicated and used. It details how assessment data can be optimised for different uses and outlines the support needed to implement the procedures.

The need for a new assessment framework

Assessment has a strong influence on curriculum and pedagogy, and it is essential to have explicit links between what is intended to be learned and what is assessed. National science testing at the end of Key Stage 2 was abolished in England in 2010 (2004 in Wales). Although assessment in science is still required, evidence suggests it is not being used effectively to help learning or to report pupils' knowledge and understanding.

The review of the National Curriculum currently taking place provides an opportunity for improving assessment policy and practice. The Department for Education published the draft National Curriculum for Key Stages 1 and 2 on 11 June 2012,[2] ahead of any draft assessment arrangements. Although the draft National Curriculum was published after the seminar, the report makes reference to some of its proposals.

1 This report, *Developing policy, principles and practice in primary school science assessment*, is also available to download from www.nuffieldfoundation.org/primary-science-assessment.

2 Department for Education, 2012. *National Curriculum for Science Key Stages 1 and 2 – Draft* [pdf]. Available at: http://media.education.gov.uk/assets/files/pdf/d/draft%20national%20curriculum%20for%20science%20key%20stages%201%202.pdf.

Principles and purposes of assessment

The report sets out explicit principles that should underpin decisions about the role of assessment in primary science education and the methods of collecting assessment data. Key principles are that all assessment should 'ultimately improve pupils' learning' and that it should be aligned with the full range of learning objectives.

The framework proposes that assessment of pupils' achievements gathered within the school is used for two main purposes:

- to help learning (formative or 'assessment for learning');
- to summarise and report on what has been learned (summative or 'assessment of learning').

For the purpose of monitoring national performance, the framework proposes the continuation of testing random samples of pupils, albeit in a modified form.

Assessment methods

There is a broad consensus that **assessment by teachers** is the most appropriate form of assessment for science in the primary school. One of the key aims of science at this stage is to enable children to 'work scientifically', something that cannot be assessed by external written tests alone. Teachers can observe pupils when engaged in science investigations and ascertain their understanding by listening, questioning and looking at their work. Scientific vocabulary and factual knowledge can be assessed efficiently by short tests or quizzes given by the teacher at appropriate times.

The proposed framework is based on the understanding that teachers should use formative assessment as part of their everyday practice to help pupils achieve learning outcomes. Information gathered as part of teaching should be used to provide short, narrative annual reports for parents and the next teacher, and summative assessments of achievement at the end of each key stage. A programme of ongoing support should be put in place to build teachers' capability and confidence in their assessment.

The proposed assessment framework adopts the key stage (KS) structure used in the draft National Curriculum: KS1 (Years 1 and 2), Lower KS2 (Years 3 and 4) and Upper KS2 (Years 5 and 6). Each key stage has a Programme of Study, which sets out the knowledge, skills and understanding that should be achieved by pupils at the end of that stage.

The proposed framework for primary school science assessment

1. Teachers use **formative assessment** throughout the year to support pupils' achievement.

2. Information gathered by teachers during the year is used to provide a **narrative annual report** to parents and is passed to the pupil's next teacher or school.

3. At the end of each key stage (KS1, Lower KS2 and Upper KS2), teachers judge whether or not each pupil has achieved the designated **learning outcomes for the Key Stage** in the main components of the National Curriculum ('science knowledge and understanding' and 'working scientifically').

   These judgments should be reported in terms of whether the relevant learning outcomes have been 'achieved' or 'not yet achieved' (an intermediate category could also be trialled to facilitate more nuanced decisions). This information should be used to show the progress that individual pupils have made as they pass through school.

   The translation of **detailed formative data** to **summative judgements** should be moderated within the school, using group procedures and reference to national exemplars.

4. At the end of each key stage, individual pupil records are aggregated for each class, for the school as a whole and for particular groups. The data are used for internal school evaluation and to track the progress of groups such as higher and lower achieving pupils.

5. Each year, data relating to the proportion of Year 6 pupils who achieve the learning objectives for Upper KS2 is reported to parents and governors and published on the school website (but is not collected centrally).

6. At a **national level**, performance is tracked via annual testing of random samples of Year 6 pupils. Several different sets of test items from a bank are used in order to employ the maximum number of items while minimising the impact of the testing. The sample test results should be the only data reported at national level.

   The sample test results should be used to publish short reports relating to particular learning outcomes. This will enable schools and parents to evaluate pupils' performance within a national context.

1. Introduction

The discussion and recommendations reported here are the outcome of a seminar held at the Nuffield Foundation on 7 and 8 June 2012. The purpose of the seminar was to discuss the assessment of science in primary schools in the context of the revision of the primary curriculum and its assessment in England.

National science testing at the end of Key Stage 2 was abolished in England in 2010 (2004 in Wales) in recognition that it failed to provide relevant and useful information about pupils' performance and had a constraining impact on teaching. However, it was not replaced by a coherent framework for assessment designed to support learning and provide dependable information for reporting achievement at pupil, school and national levels. The aim in holding the seminar was to develop such a framework. Three days after the seminar, the Department for Education published a draft National Curriculum for Key Stages 1 and 2, covering the primary school years for pupils aged 5 to 11 years. The draft did not influence the seminar discussions, but this report includes reference to some of its proposals.

The overall aim of the seminar was to set out the views of an informed group of educators about the forms that assessment should take in the coming policy changes. The participants were leaders in primary science education and members of key organisations concerned with promoting excellence in the field. Participants began by studying relevant documents in advance of the seminar. The framework was developed collaboratively during the seminar discussions and afterwards by correspondence. The final report offers an approach to assessment in primary school science education for consideration by organisations other than those represented at the seminar.

The Nuffield Foundation funded the seminar, but the views expressed are those of the participants and not necessarily those of the Foundation.

2. Background

Assessment has a strong influence on curriculum and pedagogy, making it essential that there is a correspondence between what is intended to be learned and what is assessed. The draft National Curriculum for science Key Stages 1 and 2, published on 11 June 2012, gave no indication of when draft assessment arrangements will follow. This provided an opportunity to take a fresh look at the possibilities for policy and practice in assessment of primary science in advance of responding to government proposals.

The termination of national testing in science at the end of KS2 in England in 2010 (2004 in Wales) has had a significant impact on teaching practice. Despite criticisms of the KS2 science tests and the endorsement of their removal by the Independent Review of KS2 Testing, Assessment and Accountability (the Bew Report),[3] it was apparent that the testing had served to ensure that attention and time were given to science appropriate to its status as a core subject. Various studies (from the Wellcome Trust,[4] the National Science Learning Centre and the research using the Primary Science Quality Mark data[5]) have shown that when the science tests were removed, time and attention were transferred from science to the other core subjects of mathematics and English, where high stakes testing remained. Although assessment by teachers continued to be required in science, too little attention has been given to ensuring its quality and dependability. In general, teachers have relied on test items to guide their assessment. After national testing was abolished, teachers continued to use past test papers or published tests rather than the rich source of data that classroom science activities afford. Thus the potential for assessment by teachers to record valid information on important outcomes that tests cannot provide, and more authentic data on aspects of learning that tests do assess, has not been realised.

One attempt to support assessment by teachers was the now abandoned Assessing Pupils' Progress (APP) project. APP endeavoured to provide a structure and help for teachers to make judgements about the National Curriculum levels achieved by their pupils. It embodied the idea of using criteria to judge the level of pupils' performance at particular times and to identify progress. The project provided examples of pupils' work to illustrate the meaning of the criteria. However, the APP 'assessment focuses' were not aligned to the National Curriculum and did not provide for reporting across the full range of learning outcomes. Consequently the procedures were additional to, rather than integral to, the requirements of the National Curriculum Assessment. The procedures were also found to make heavy demands on teachers' time if they were to be used for all pupils.

The indications were that science was being marginalised and, importantly, that assessment was not being used effectively either to help learning or to report the

3 Lord Bew, 2011. *Independent Review of Key Stage 2 testing, assessment and accountability: Final Report* [pdf]. Available at: https://media.education.gov.uk/MediaFiles/C/C/0/%7BCC021195-3870-40B7-AC0B-66004C329F1F%7DIndependent%20review%20of%20KS2%20testing,%20final%20report.pdf.

4 Welcome Trust, 2010. *Attitudes of Children and Parents to Key Stage 2 Science Testing and Assessment.* Also: Collins, S., Reiss, M & Stobart, G. (2008) *The Effects of National Testing in Science at KS2 in England and Wales*. Commissioned by the Wellcome Trust in association with the ASE.

5 Turner, J., Marshall, S. and Farley, A. In press. *What can we learn about current practice in primary science teaching and learning in England from the primary science quality mark award programme?*

knowledge, understanding, competences and attitudes that science aims to help pupils develop. This clearly required some action from concerned primary science educators and the seminar reported here is the first step in that process. Ideally it would be followed by a sustained period of development and trial of approaches and methods of assessment. Trials should be evaluated in terms of the necessary requirements of validity, reliability, usability, efficiency and effectiveness.

3. Aims of the seminar

The overall purpose of the seminar was to propose a framework for assessment of primary pupils' attainment in science. An assessment framework sets out how evidence of pupils' attainment should be collected, recorded, communicated and used by those involved in pupils' education. It describes how the dependability of the resulting information can be optimised for different purposes and what support is needed to implement the procedures. It is particularly important to ensure that the procedures enable assessment to serve its two main purposes in the primary school: to help learning (formative or 'assessment for learning') and to summarise and report on what has been learned (summative or 'assessment of learning'). In the secondary years assessment is also used for pupil selection and for certification of attainment.

What determines whether assessment is formative or summative is not what information is collected or how it is collected but how it is *used*. The principal use of formative assessment is to help learning. It provides 'evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how to get there'[6]. The Bew Report identified the principal use of summative assessment as being 'to ascertain what pupils have achieved in relation to the attainment targets for that stage'[7]. It also noted three other uses: holding schools accountable for the standard of attainment and progress of their pupils; informing parents and secondary schools; and enabling benchmarking between schools. These are only some of the many other uses of summative (particularly test) data as identified by Newton,[8] not all of which are regarded as appropriate.

6 Assessment Reform Group (ARG) 2002 *Assessment for Learning: 10 Principles* [pdf]. Available at: http://assessmentreformgroup.files.wordpress.com/2012/01/10principles_english.pdf.

7 *Independent Review of Key Stage 2 testing, assessment and accountability: Final Report* (Bew report), p.17.

8 Newton, P.E. 2012. Validity, purpose and the recycling of results from educational assessments. In J. Gardner, ed. *Assessment and Learning*, 2nd edition. London: Sage.

The importance of formative assessment to learning is now widely supported by research evidence[9]. It is also recognised that the distinction between formative and summative assessment is not clear-cut. Not all information that could be used to help learning is indeed used for this purpose and some information can be used for both formative and summative purposes. Thus a framework for assessment that promotes pupil progress and is aligned with the curriculum should include reference to how evidence of learning is used formatively as well as summatively.

In summary, the following aims were identified for the seminar:

1. To set out explicit principles that should underpin the assessment of science in the primary school.
2. To provide a framework, underpinned by a rationale, for assessment and its use in teaching and learning science as children progress through the primary school.
3. To identify ways in which the various goals of learning science can be assessed.
4. To consider how achievement in science may be recorded and reported.
5. To suggest ways in which schools can be supported to conduct both formative and summative assessment, drawing on research evidence and existing procedures.

Arguments and proposals in relation to these aims are set out in the following sections.

4. Principles for primary science assessment

Statements about the role, aims and processes of education are unavoidably based on views of what is considered worthwhile and valuable to learn. This is especially true of decisions about assessment, which always involve the selection of what to assess and how to do it. What is actually assessed is inevitably a sample of what could potentially be assessed. A different sample would produce a different result. So the outcome of assessment is only ever an estimate, not an exact measurement or a statement of everything a child knows and can do. Consequently, it is important to make explicit the values and standards which have guided the proposals in this report. This is the purpose of the following statement of principles.

9 OECD, 2005. Formative Assessment: Improving Learning in Secondary Classrooms [pdf]. Available at: http://www.oecd.org/dataoecd/19/31/35661078.pdf.

In relation to its functions, the group considered that assessment should:

- ultimately improve pupils' learning
- be aligned with the full range of learning objectives of the whole school curriculum[10]
- be an integral part of teaching that enables pupils to understand the purpose of their activities and to improve the quality of their work
- combine qualitative and quantitative data of different kinds, from a variety of individual and group learning activities including pupils' self-assessment, to inform decisions about pupils' learning and achievements
- be understood as providing data about pupils' learning outcomes in the form of approximations and samples, subject to unavoidable variations.

In relation to the methods and procedures of assessment, the group considered that these should:

- promote the active engagement of pupils in their learning and its assessment, enabling and motivating them to show what they know and can do
- include explicit processes to ensure that information is valid, reflecting all important learning goals, and is as reliable as necessary for its purpose
- meet standards that reflect a broad consensus on quality at all levels from classroom practice to national policy
- be realistic and manageable for pupils and teachers, with transparent time demands and requiring no more collection of pupils' work than is a normal part of teaching and learning.

5. The challenge of assessing primary science

Unlike English and mathematics, science has not always been regarded as an essential part of primary education. The reasons for its gradual introduction during the 1960s and 1970s, its incorporation into government policy and eventual establishment as a 'core' subject of the National Curriculum have been well documented. The main research-based arguments for beginning science in the primary school have been summarised as follows:

10 As proposed by the Expert Panel, the National Curriculum is taken to be the statutory part of a wider school curriculum as experienced by the pupils. Reference to the statutory part here is indicated by capital letters, using a small 'c' when referring to the wider curriculum.

> Learning science helps children to begin to understand aspects of the world around them, both the natural environment and that created through the application of science. Primary science gives children the opportunity to begin forming key concepts through exploring a variety of contexts before formal learning at the secondary level. The evidence that children are arriving at their own ideas in the early years, whether or not there is science in the curriculum, is a powerful argument for ensuring that they explore and inquire in a way that promotes the development of reliable knowledge and basic science concepts. There is a considerable body of research evidence that shows that, since children's own ideas are often in conflict with scientific ones, if taken into the secondary school, these can inhibit effective learning. The conflict between children's own ideas and ones that they are taught in secondary education leads many to find science too hard, too confusing and too remote from their real experience.[11]

The challenge of providing valid assessment data about pupils' learning in science becomes apparent by considering the aims of primary science education, usually expressed in terms of developing:

- concepts that help pupils to understand the world around them
- scientific capabilities concerned with gathering and using evidence
- scientific attitudes
- a language for discussing, describing and communicating their observations, ideas and conclusions.

In the draft National Curriculum for England published on 11 June 2012 the aims are stated as ensuring that all pupils:

- develop scientific knowledge and conceptual understanding through the specific disciplines of biology, chemistry and physics
- develop understanding of the nature, processes and methods of science through practical activity
- are equipped with the scientific knowledge required to understand its uses and implications today and in the future.[12]

The draft National Curriculum for Science Key Stages 1 and 2 proposes Programmes of Study in relation to knowledge and understanding and 'working

11 Harlen, W. (2008) Science as a key component of the primary curriculum: a rationale with policy implications. *Perspectives on Education 1 (Primary Science)*, 2008 4-18. Available at: www.wellcome.ac.uk/perspectives.

12 Department for Education, 2012. *National Curriculum for Science Key Stages 1 and 2 – Draft* [pdf], p. 1.

scientifically' for three key stages: KS1 (Years 1 and 2), Lower KS2 (Years 3 and 4) and Upper KS2 (Years 5 and 6). This key stage structure has been adopted in the assessment framework proposed here.

The attainment targets for the key stages are the achievement of the knowledge, skills and understanding specified in the programme of study. The draft National Curriculum is used as an example to show how the assessment framework can be aligned with the intended learning outcomes, while not endorsing the draft Programmes of Study. Indeed, the group would favour a Programme of Study which sets out expected learning outcomes for each key stage as a progression toward the development of big ideas of science and ability to work scientifically. However, it was not the purpose of the seminar to comment on the content of the curriculum.

6. Methods of assessment

Assessment is the process of collecting, analysing and interpreting evidence and using it to make inferences about what pupils know and can do. In order for pupils to show their competence in 'working scientifically', they need to be in situations where they are raising questions, planning investigations, observing, measuring, analysing, arguing and evaluating. They must also be engaged in supporting their conclusions with argument and evidence and in working with and learning from others. Such situations also provide opportunity for pupils to develop and demonstrate attitudes relating to willingness to take account of evidence, which are not made explicit in the draft National Curriculum. To be able to show their knowledge and understanding, pupils should be required to apply their ideas to new events and phenomena.

Collecting evidence of knowledge, understanding, skills and attitudes involves a range of different assessment methods. The ability to 'work scientifically' can only be validly assessed when pupils are in situations where scientific work such as planning and carrying out investigations is taking place. The application of knowledge is best assessed through discussion of events and phenomena that are new to the pupils. Other aims of primary science education, such as knowledge of scientific vocabulary and procedures, can be assessed efficiently by short tests or quizzes administered at appropriate times by the teacher. But it is not possible for all the aims of primary science education to be validly assessed through external written tests, such as the national tests, a point acknowledged in the Bew Report. These arguments lead to the conclusion that the situations in which children learn science provide the best opportunities for assessment of their learning. This point was accepted in the government's response to the Bew Report which expressed the view that 'teacher assessment is the most appropriate form of assessment for

science at the end of Key Stage 2, so pupil and school level data will continue to be based on teacher assessment judgements'.[13]

For monitoring national performance in science, the government accepted the Bew Report recommendation to continue to use tests administered to a sample of pupils at the end of KS2. In this context the test results are not reported for individual pupils, but nevertheless the validity of the assessment is essential to their usefulness. If the test is restricted to the number of items that an individual pupil can reasonably be expected to answer in a limited time, the test is unlikely to capture performance across the range of aims of the curriculum. The reliability of the results is limited by the inevitability, mentioned earlier, of including only a small part of the range of potential items. However, in sample testing it is not necessary for all pupils to be given the same items. A large number of questions, sampling different skills, concepts and contexts, can be used. The limitations on the range of performance that can be assessed using written tests will remain. But the further limitation due to using only a small number of items can be reduced, providing that the widest possible range of items is used and not the single tests that were used in 2010 and 2011. Thus, although sample testing does not provide data about all expected learning outcomes, it has a useful role in national monitoring (as it does in international surveys such as PISA and TIMSS, where the scores of individual pupils are not relevant).

Previous experience of national monitoring, by the Assessment of Performance Unit in the 1980s, showed the value of using a large bank of assessment items. Results provided information not only about whether overall attainment was changing, but also about performance in different aspects of subjects, and among sub-groups of pupils in different learning conditions. Short reports for teachers, which gave examples of responses to groups of items, also enabled all schools, whether included in the sample of not, to benchmark the performance of their 11-year-old pupils in relation to particular skills or conceptual learning. In future surveys, such brief reports, with examples of items and how pupils responded to them, ought to be part of the information widely available as feedback to parents as well as teachers. Knowing the areas of strength and weakness in national performance affords opportunities to reflect on the achievement of their pupils in these areas.

13 Department for Education, 2011. *Independent Review of Key Stage 2 testing, assessment and accountability: Government Response* [pdf], p.10. Available at: https://www.education.gov.uk/publications/eOrderingDownload/CM-8144.pdf.

7. Recording and reporting attainment in science

The concept of identifying progression in learning in terms of levels of achievement was introduced by the Task Group on Assessment and Testing in its 1987 report. The report proposed that what pupils have learned would be measured against national standards. It argued that although this learning could be reported in the same descriptive terms as the attainment targets, it was 'advantageous also to convert the descriptions into a marking scale'. Two reasons were given for this: 'One is to convey some sense of where a child is in the process of learning; the other is to make easier the analysis of results from groups of children'.[14] So the subject working groups on the curriculum were required to identify progression across the compulsory years of school (from age 5 to 16) in 10 levels (reduced to eight in 1995). Thus for the past 22 years the attainment of pupils has been described in terms of levels, and the percentages of pupils reaching certain target levels at particular ages has been used to measure performance of individual pupils, schools, local authorities and the nation as a whole. During that time there has been increasing concern about the use of levels. Some of these concerns were expressed in evidence to the Bew Report and to the Expert Panel for the Curriculum. The latter expressed the view that the ways in which levels are currently used 'may actually inhibit the overall performance of our system and undermine learning.'[15]

Not only has the language of levels been used inappropriately to label pupils but the proliferation of sub-levels (4a, 4b, 4c, and so on) has introduced demarcations with very uncertain meanings, for which there is little evidence in the cognitive development of pupils. Moreover, it is not only parents who are unsure of what a level means in terms of what their child can do. Teachers are not able to make sense of levels provided by national test data without looking at pupils' responses to particular items.[16] Having turned descriptions of behaviour into levels in order to use the information, it now seems necessary to turn levels back into the detail of what pupils can and cannot do. Consequently it seems sensible to consider

14 Task Group on Assessment and Testing, 1987. *A Report*. London: DES and Welsh Office, Para 96.

15 Department for Education, (2011). *The Framework for the National Curriculum. A report by the Expert Panel for the National Curriculum review* [pdf], p.44. Available at: https://www.education.gov.uk/publications/eOrderingDownload/NCR- Expert%20Panel%20Report.pdf.

16 Oates, T. (2010) *Could do better. Using international comparisons to refine the National Curriculum in England.* Cambridge: Cambridge Assessment. Available at: http://www.cambridgeassessment.org.uk/ca/digitalAssets/188853_Could_do_better_FINAL_inc_foreword.pdf.

an alternative that bypasses the use of levels and reports what pupils can do in descriptive terms.

The form of reporting has to take into account the use and the users of the information. The main distinction is between information about individual pupils and information about groups or cohorts of pupils.

Individual pupil assessment and records

For ascertaining and reporting the achievement of individual pupils, information can be gathered by teachers in a variety of ways. As suggested earlier, these include observation, discussion, short teacher-made tests and pupils' written and oral reports. This information can be used to produce a narrative overview of areas of success and areas where the pupil requires further support, relating to the whole science curriculum as planned by the school. Reporting in this way is already the practice in many schools for English and mathematics. Information in this form can be conveyed to parents in the annual report and to the pupils' next teacher or school.

Schools also need to keep end of key stage records on the performance of individual pupils in the main components of the National Curriculum. This requires decisions about the nature and number of components for reporting. One solution would be to use four components, such as the attainment targets Sc1, Sc2, Sc3 and Sc4 in the current National Curriculum. A simpler version would be to reduce these to only two components, such as 'science knowledge and understanding' and 'working scientifically'. In either case the procedure would be to compare the evidence of a pupil's achievement with the expected learning outcomes for the key stage (i.e. the knowledge, skills and understanding to be taught in the key stage).

The results might be expressed as whether outcomes have been 'achieved' or 'not yet achieved'. Alternatively, judgements might be in terms of the learning outcomes being 'mastered' or 'not yet mastered', or 'ready to progress' or 'not yet ready to progress'. There is also the option to use more than a binary distinction, inserting an intermediate between 'achieved' and 'not yet achieved'. Each approach has its pros and cons which could be explored through trials. In the framework proposed in this report, we have used 'achieved' and 'not yet achieved' as being most straightforward and avoiding the creation of distinctions that would share the deficiencies of sub-levels. However the result is expressed, the judgement needs to be as trustworthy as possible, since it will contribute to the information used internally and externally in evaluating teachers and the school. Hence some form of moderation will be needed at the end of each key stage. This is considered in section 8.

Aggregated records for classes and the school

Uses of assessment relating to accountability, evaluation, and benchmarking require the aggregation of data for groups of pupils at the end of key stages and particularly at the end of KS2. This provides schools with records of the proportion of pupils who achieve the learning outcomes for each key stage. The results for each pupil can also be aggregated in different ways and used to study the achievement of particular groups of pupils, perhaps by gender, background or achievement in other core subjects. Results can also be used to trace the achievement of the lowest and highest achieving pupils as they move through the school, thus enabling the school to show, for example, that it is helping those who may not have achieved the outcome at the end of KS1 to achieve the intended learning in later key stages.

8. Supporting teachers' assessment

Teachers have been required to assess pupils' achievement since the introduction of National Curriculum Assessment in 1988, but many teachers lack confidence and competence in carrying out this part of their work. Some reasons for this have been revealed in research studies of teachers' assessment in England.[17] The curriculum and the assessment arrangements were introduced first at KS1. Early studies found that the introduction of teachers' assessment initially had a beneficial effect on Year 2 teachers' planning and was integrated into teaching. This was particularly the case where teachers worked collaboratively, sharing their interpretation of the level descriptions and their judgements of examples of pupils' work. Subsequently however, in the later 1990s, as support for teachers' assessment was removed, there was a decline in earlier collaboration among teachers and shared interpretations of criteria. Lack of funding for teachers to moderate their assessment results and for professional development in assessment, together with the priority given to test results, conveyed the message to teachers that their assessment of pupils was not valued.

It is important to learn from this experience if the benefits for validity of assessment in science are to be gained. Steps need to be taken – and shown to be taken – to improve the reliability of teachers' assessment, which is necessary for confidence in the results and for preserving validity. The most commonly used methods for assuring the quality and reliability of assessment by teachers are:

17 Harlen, W. (2005) Trusting teachers' judgement: research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research Papers in Education* vol 20 (3) 245-270.

using exemplars, group moderation, and using some form of test or special task as a check of teachers' judgements but not to be reported as a part of the result.

A national collection of exemplars of pupils' work, sometimes annotated to highlight aspects that are significant evidence of achievement, can enable teachers to adjust their understanding of the criteria and how they are to be applied. Since teachers should be basing their summative judgements on a range of each pupil's work, the most useful exemplar material is in the form of a portfolio. It is not expected that teachers will find an exact match between exemplars and their pupils' work; the role of the exemplars is to communicate the essential features of successful achievement at the end of each key stage. Where possible, the exemplars should cover the types of evidence that teachers will use in their assessment – pupils' talk, actions, writing, drawings and artefacts.

In group moderation teachers meet to discuss examples of pupils' work and align their understanding and application of the criteria for achievement. This has benefits besides improving the quality of teachers' assessment of pupils. Meeting to discuss the inferences that can be drawn from studying pupils' work provides insights into the assessment process and also leads to discussion of how to support pupils' progress – that is, how to use the information about a pupil formatively as well as summatively. The process should be seen as quality assurance rather than quality control, using examples of work normally available and not requiring the special collection of evidence.

The use of a brief test as a means of moderating or checking teachers' judgement is used in primary schools in Scotland for mathematics and reading and writing, but no longer in science. The test results are not a separate measure of achievement but only used when teachers judge a pupil to have reached a certain level. They are not a requirement, and teachers can choose to use the alternative of group moderation of their judgements. As in the case of the use of exemplars, this approach can be used in circumstances where teachers may not be able to meet for moderation purposes, but lacks the benefits of group moderation in terms of professional development.

Another essential resource to help teachers develop their assessment practice is time. Once established, both formative assessment and summative assessment enable teachers to make better use of teaching time. In the stages of developing practice, however, time to attend professional development in assessment, either in-school or at a centre, will be needed. Teachers may also need to spend more time providing reports on each pupil at the end of every year. In terms of helping pupils learn, this is a more productive use of time than preparing for and administering national tests.

9. The proposed assessment framework

The discussion of methods of assessment and forms of reporting in sections 6 and 7 leads to an assessment framework which separates data for use within the school from data for use at the national level. Data used for different purposes within the school are collected by teachers. Data about national standards of performance are collected by entirely separate procedures involving sample testing.

Within the school, teachers gather data about individual pupils and groups of pupils as part of teaching and use it to inform their teaching and to help learning. This data collection is depicted as the base of a pyramid in **Figure 1**. At the end of each year these data are used to provide a narrative overview reported to parents and passed to the pupils' next teacher or school. At the end of each key stage, teachers' judgements, suitably moderated, are used to provide summative assessment of achievement in terms of a small number of components (preferably two) of the National Curriculum in Science. Thus there is a gradual reduction in the breadth and detail of information that is recorded and reported, from the rich formative assessment to the succinct, summative information.

For use at class and school level, there is a further reduction as the data are passed up the pyramid. Individual pupil data are aggregated to give class and school numbers or percentages of pupils judged as having achieved the learning outcomes of each key stage. These data are used by the senior management for internal school evaluation and fed back to classes to be used in reviewing and improving performance. The percentages of pupils achieving the learning outcomes at the end of Upper KS2 is a particularly important part of the account of the school's performance, included in the Annual Report and published on the school's website. It may also be part of the information required by inspectors. It is proposed that end-of -Upper-KS2 data should not be locally or nationally collected, to deter the creation of league tables based solely on this one aspect of performance. While recognising that comparisons between schools cannot be entirely avoided when reports are in the public domain, presenting specific information on achievement in the context of the school's broader performance should encourage responsible use of the data.

For reporting on performance across the country, it is important for the information to be based on performance across the full range of learning outcomes for science in the National Curriculum. The use of a single test given to all pupils has been shown to have widely damaging effects,[18] and to provide data

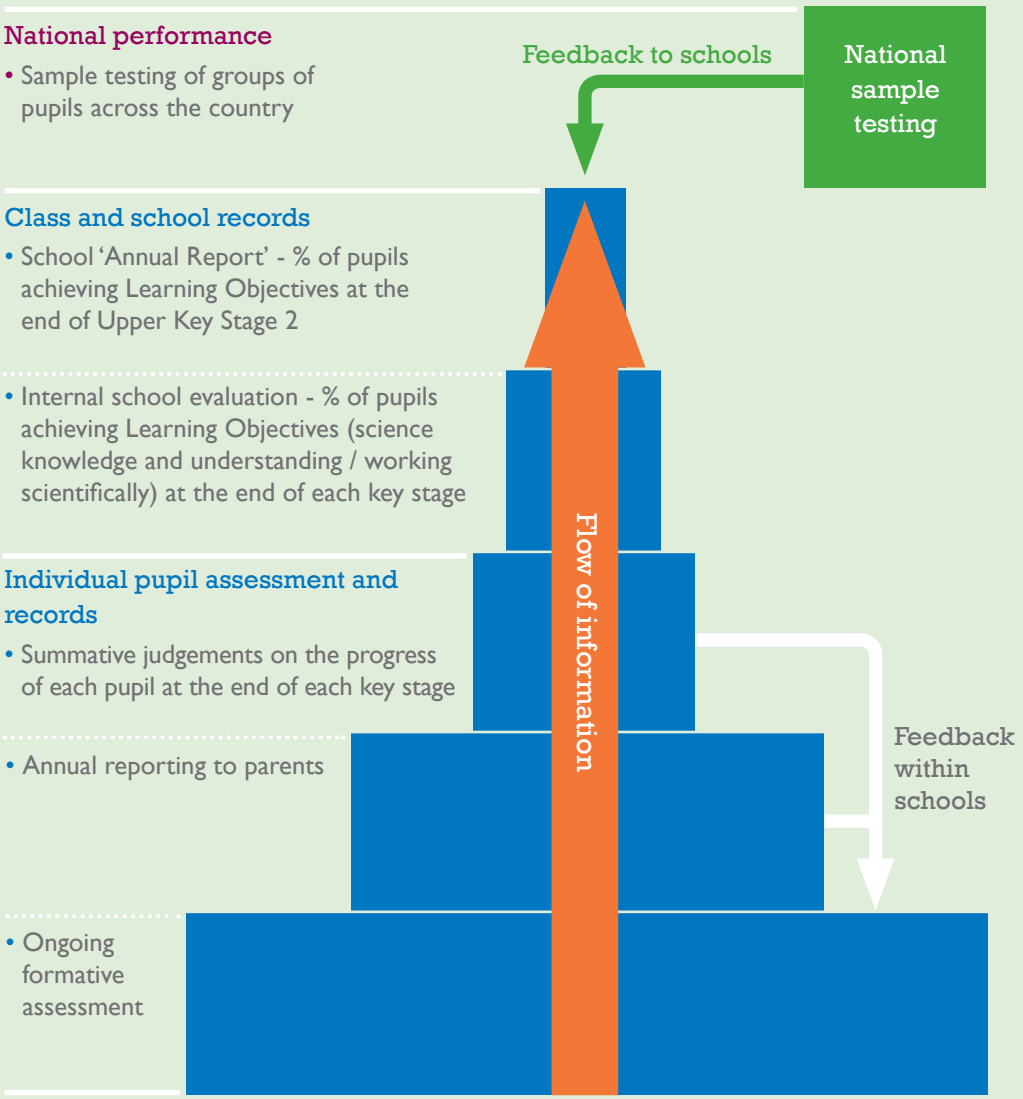
18 Harlen, W and Deakin Crick, R. (2003). Testing and Motivation for Learning, *Assessment in Education*, vol 10 (2) 169-208.

of limited dependability.[19] The dependability can be increased by using a number of items that among them cover a variety of concepts and content. It is proposed, therefore, that national performance should be assessed using a bank of items and a design in which only a small proportion of pupils take part. Each pupil responds to a sample of the full range of assessment items. Results for individual pupils and schools in the sample have no meaning until combined with others, discouraging 'teaching to the test'. Annual reports can show national trends in performance across years in different skills and areas of understanding. The reports should enable all schools, whether or not in the sample, to reflect on the teaching and learning of their Year 6 pupils.

Figure 1 shows the framework presented as a diagram. The type and use of data is summarised in **Table 1**.

19 Newton, P. (2009). The reliability of results from national curriculum testing in England, *Educational Research*, vol 51(2) 181-212.

Figure 1: The flow of assessment data through the school



Individual, class and school assessment: A wide range of assessment processes provide a rich and extensive source of information on individual pupil's learning, which is used to report on achievement. Information on individual pupils is then summarised to provide summative judgements which can be collated at the level of class or whole school. Class and school level data can be used for internal reporting and the School Annual Report. Feedback from internal evaluation and summative judgements informs future improvement of teaching and learning.

National performance: Sample testing of groups of pupils across the country provide data on the national picture which in turn provides feedback to schools.

Table 1: Assessment data and use at pupil, school and national levels

Time	Data	Use
Individual pupil assessment and records		
Throughout each year	Teachers gather data about each pupil as part of their teaching of the whole science curriculum for that year.	Throughout each key stage, teachers use formative assessment to ensure the progress of all pupils in the whole science curriculum of the school, including the learning set out for the key stage. At the end of each year, teachers provide a narrative report to parents/carers giving an overview of their child's achievement and what is still to be achieved. This information is passed to the next teacher or school.
At the end of KS1, end of Lower KS2 and end of Upper KS2	Teachers summarise the attainment of each pupil, by judging whether or not each pupil has achieved the overall learning outcomes for the key stage in the main components of the National Curriculum ('science knowledge and understanding' and 'working scientifically'). These judgements are moderated internally within the school.	The records of individual pupils at the end of each key stage are used to show the progress that individual pupils have made as they pass through the school.

Individual pupil assessment data used to compile class and school records

Class and school records		
At the end of each key stage	Individual pupil records are aggregated for each class and the school. This gives numbers (or percentages) of pupils having achieved the learning outcomes in 'science knowledge and understanding' and 'working scientifically'.	Records of the attainment of pupils in each class and the whole school are used for internal school evaluation by the senior management team. Proportions of pupils achieving the expected learning outcomes at the end of each key stage provide evidence of the progress made by groups of pupils, such as those who may not have been able to achieve the expected learning at the end of KS1 but were able to do so for later key stages.
End of Upper KS2	The proportion of pupils in Year 6 assessed each year as having achieved the learning objectives for Upper KS2.	End of Upper KS2 results reported to parents and governors, and published on the school website as part of the public information about the school for external evaluation. (No central collection of data.)

National performance		
Annually for a sample of pupils in Year 6	Tests drawn from a bank of test items are administered to equivalent, random samples of Year 6 pupils. Using several sets of items extends the range of items included in the surveys, while minimising the impact of the testing. National performance is reported each year relating to aspects of 'science knowledge and understanding' and 'working scientifically'.	Provides a national picture, enabling year-on-year performance to be tracked. Reports on aspects of 'science knowledge and understanding' and 'working scientifically' enable schools to reflect on their performance against the national picture.

10. Conclusion and recommendations

The proposed framework for assessment of science in primary schools is founded on the principles set out in section 4. It provides for the use of assessment to improve learning and for the involvement of pupils through formative assessment. Information based on high quality assessment by teachers is used within the school for reporting on outcomes of the National Curriculum and the wider school curriculum. By making use of data available to teachers through their normal work, the procedures become part of their professional activities.

The arguments and evidence in this report lead to the following recommendations for a national assessment framework:

1. Teachers should use formative assessment as part of their everyday practice to help pupils achieve the learning that is set out in the school's curriculum for science.

2. The rich description of each pupil's achievement, provided through formative assessment, should be used for a narrative annual report to parents on the learning outcomes for the year. The report should indicate where help is being given to ensure that the expected learning can be achieved by the end of the key stage.

3. This description of each pupil's performance in relation to the learning outcomes for each year should be passed on to the next teacher or school.

4. Assessment should be based on teachers' judgements so that a range of outcomes relating to knowledge, understanding, skills and attitudes can be assessed.

5. Information gathered and used for formative assessment and for reporting to parents should be used to produce the summative assessment at the end of each key stage for the two key components of the science curriculum.

6. Attainment should be summarised in terms of achievement of the learning objectives for the key stage. The resulting judgements at the end of each key stage should be reported as the learning outcomes being 'achieved' or 'not yet achieved'. An intermediate category could also be trialled to facilitate more nuanced decisions.

7. The move from detailed formative data to summative judgements at the end of each key stage should be subject to moderation within the school, using group procedures and reference to national exemplars.

8. A programme of ongoing support should be put in place to build teachers' capability and confidence in their assessment of pupils' learning, as part of developing understanding that assessment is an integral part of effective teaching.

9. At a national level, performance in science should be tracked through annual testing of samples of pupils in Year 6. The tests should be created from a bank of items, with different samples of pupils taking different tests thus extending the range of items that are used in the surveys and the validity of the reported results.

10. The sample test results should be the only data to be reported at the national level. Teachers' assessments should not be centrally collected.

11. Short reports should be published on the national performance on groups of items used in the sample test relating to particular learning outcomes. This will enable schools to compare their Year 6 pupils' performance with the national picture and allow parents to reflect on their child's learning in relation to national performance.

Published by the **Nuffield Foundation**, 28 Bedford Square, London WC1B 3JS
Telephone +44 (0)20 7631 0566
www.nuffieldfoundation.org