Approaches to Science Assessment in English Primary Schools

Interim findings from the Teacher Assessment in Primary Science (TAPS) project

Dan Davies, Christopher Collier, Sarah Earle, Alan Howe and Kendra McMahon
Centre for Research in Early Scientific Learning (CRESL), Bath Spa University

May 2014
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOREWORD</td>
<td>3</td>
</tr>
<tr>
<td>BACKGROUND</td>
<td>5</td>
</tr>
<tr>
<td>METHODOLOGY</td>
<td>6</td>
</tr>
<tr>
<td>BROAD APPROACHES TO FORMATIVE AND SUMMATIVE ASSESSMENT</td>
<td>7</td>
</tr>
<tr>
<td>CASE STUDIES OF TYPICAL APPROACHES TO SCIENCE ASSESSMENT</td>
<td>8</td>
</tr>
<tr>
<td>SUMMARY &amp; CONCLUSIONS</td>
<td>19</td>
</tr>
<tr>
<td>NEXT STEPS</td>
<td>21</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>22</td>
</tr>
</tbody>
</table>
“There would be no need for assessment if students understood all that was taught.”

According to this extreme position, a simple record of the elements which had been taught would suffice, but (for many good reasons) this is not the case. So some form of assessment is clearly necessary; but what is appropriate, in terms of assessment type and frequency, can tax even the most experienced teacher.

Following on from a project on Assessment in Primary School Science funded by the Nuffield Foundation and led by Wynne Harlen, the Primary Science Teaching Trust has funded Bath Spa University, through their Centre for Research in Early Scientific Learning (CRESL), to build on this initial work. The three-year project, called TAPS (Teacher Assessment in Primary Science), will survey a wide range of teachers and their schools to look at assessment practices, and will synthesise these data so that best practice can be shared and celebrated, providing a general framework for science assessment in primary schools.

Preliminary analysis of a group of Fellows from the Trust’s Primary Science Teacher College reveals that these outstanding Primary Science Teachers have developed an array of excellent assessment tools – and such achievements are not restricted to College Fellows. Although the early indications are that there is much excellent practice being undertaken, the task of capturing and teasing out the key elements, and disseminating these ideas across the UK Primary School community, will be a significant one.

I and the Trust are confident that the Bath Spa University team will rise to the challenge and that this Primary Science Teaching Trust Hub will produce materials that will be of immense value to the primary school science teaching community.

Prof. Dudley E. Shallcross
Director of the Primary Science Teaching Trust
This report aims to survey and categorise the different [assessment] approaches currently being undertaken by schools, whilst reviewing their strengths and weaknesses.
Since the discontinuation of Standard Attainment Tests (SATs) in science at ages 7 and 11, pupil performance data in science (expressed as National Curriculum levels) reported to the Department for Education by each primary school in England has relied largely on teacher assessment (TA) undertaken in the classroom. Subject leaders’ submissions to the Primary Science Quality Mark (PSQM) database indicate that most regard assessment as a problem to be solved; with many mentioning a lack of consistency across the school and a need for staff training.

Schools are trialling a variety of different methods – including Assessing Pupil Progress (APP), commercial schemes and ‘home-grown’ solutions – with many subject leaders expressing the anxiety that any single approach on its own is insufficient. The majority are planning to moderate and continue to review their assessment systems (Earle 2014). This report aims to survey and categorise the different approaches currently being undertaken by schools whilst reviewing their strengths and weaknesses.

The current review of the primary curriculum (DfE 2013) indicates that from 2014 children’s attainment in science will be summarised in terms of achievement of the learning objectives for each year group (expressed as objectives within a programme of study), and at the end of three Key Stages (ages 7, 9 and 11). It would appear that teachers’ judgements at the end of each Key Stage are likely to be reported as the learning outcomes being ‘achieved’ or ‘not yet achieved’. An intermediate category could also be trialled to facilitate more nuanced decisions.

This represents a radical shift away from the assignment of children to National Curriculum levels, regardless of age, which has been a feature of school assessment in England since 1988. Our experience of working closely with 12 schools on the TAPS project, together with presentations made to groups of teachers at Primary Science Teaching Trust (PSTT) and Association for Science Education conferences, is that very few schools have yet adapted their assessment approaches to the ‘post-levels’ world, and that most propose to continue levelling pupils during 2014-15 ‘just in case...’

A working group of experts convened by the Nuffield Foundation (2012), considering the development of policy, principles and practice in primary school science assessment, made a number of recommendations which appear to be consistent with the approach taken in the curriculum review. Amongst these recommendations are that the rich formative assessment data collected by teachers in the course of ongoing classroom work in science should also be made to serve a summative purpose (reporting to parents, teachers of the following age group, government) through synopsis at the end of academic years or Key Stages.

BACKGROUND

The Teacher Assessment in Primary Science (TAPS) project

The TAPS project aims to operationalise the Nuffield working group recommendations and develop a solution for teacher assessment of conceptual and procedural development in science which is pedagogically useful, valid, reliable and manageable in primary classrooms. The project aims to address the following research questions:

RQ1. What approaches are primary teachers currently using to assess children’s learning in science?

RQ2. How valid, reliable and manageable are these approaches?

RQ3. Can an approach be synthesised from existing good practice and ongoing development over the course of the project which meets the requirements of the revised National Curriculum; implements Nuffield recommendations; and which is valid, reliable and manageable for teachers?

RQ4. What is the potential role for ICT in enhancing validity, reliability and manageability of teacher assessment in primary science?

RQ5. What model(s) of CPD can support teachers in developing their skills to make valid and reliable assessment judgements in science whilst retaining manageability?

This report addresses research questions 1 and 2.

We have recruited 12 project primary schools from 50 applications across six local authorities in the South West of England (Bristol, South Gloucestershire, Wiltshire, Swindon, North Somerset, Somerset), selected to represent a range of urban, suburban and rural settings. Each applicant school was required to describe their current approaches to science assessment, so we were able to select a group which represented the range of approaches taken across the 50 applicants. Further comparison of these approaches with the analysis of 91 submissions to the PSQM database (Earle 2014) and discussions with Fellows and Members of the Primary Science Teacher College assure us that the range of approaches to science assessment across our 12 project schools is reasonably representative of the picture across England.
The findings reported below are based on analysis of two principal data sources:

1. The submissions to an online database of science subject leaders in all 91 English primary schools who worked towards the PSQM in Round 4 (April 2012 to March 2013). Data consist of written reflections in Spring 2013 regarding current school practice in science and developments over the past year.

2. Visits to TAPS project schools undertaken in November 2013, January and March 2014, involving interviews with science, assessment and ICT coordinators, observations of science lessons from Years 1 to 6, collection of school science and assessment policies, collection of examples of assessment tools, annotated pupil work, tracking grids, reports to parents etc.

These two data sources were treated separately. Source 1 involved content analysis of a pre-existing dataset. In order to build a quantitative picture of the types of science assessment being used by the 91 schools their written reflections were coded using Atlas.TI qualitative analysis software. To distinguish between approaches to formative and summative assessment, it was important to clearly identify a practical definition of ‘summative’ which could be applied consistently to this data set.

The method was classified as summaive if it:
- was described as ‘end of unit’ or ‘end of year’;
- fulfilled a summarising purpose, e.g. passed onto the next teacher or put into the school tracking software;
- was identified by the teacher as ‘summative’.

Formative assessment was harder to classify, partly due to the wide range of methods being employed. There is also the question of whether the strategies described were being used to identify the next steps for the learner. For the purposes of comparing methods – whether or not they were explicitly identified as supporting learning - they were termed ‘elicitation strategies’ (Harlen and Osborne 1985, Ollerenshaw and Ritchie 1997).

The wide range of elicitation strategies described across the 91 schools led to consideration of how to categorise them. Following Wiliam and Black (1996), the analysis attempted to separate the collection of assessment evidence from teacher judgement, an important consideration if exploring the possibility of using the information gathered for both formative and summative purposes.

Some elicitation strategies were classified as primarily judgemental, such as teacher marking or annotating work, self or peer evaluation. Observation and questioning were harder to classify, since it could be argued that they both involve collecting rather than judging evidence. But in recording the observation (e.g. by note taking on sticky notes or photographing) or deciding what question to ask next, the teacher is inevitably making a selection, which involves a judgement about the child’s learning and, in the case of questioning, potentially intervening. Since the mention of these techniques in a science subject leader’s summary is insufficient to separate the two purposes, they have both been included in the elicitation data for completeness.

Data from source 2 were synthesised into case studies to represent the range of approaches to teacher assessment we encountered during our visits. We used the set of indicators of effective assessment developed by Harlen (2013), which we have incorporated into the pyramid model proposed by the Nuffield group (2012) (see appendix) to make judgements concerning the validity, reliability, manageability and impact of the strategies reported by each school.
Despite the recommendations of the Nuffield group (2012) there is little evidence that the 91 schools represented in the PSQM survey are making effective links between formative and summative assessment in science (Earle 2014), since they describe the strategies used for each entirely separately (see figures 1 and 2).

From figure 1, effective features of schools’ use of formative strategies include:

- The wide range of strategies being used by schools, including talk-based (coded green), activity-based (red), observation-based (yellow) and written (blue).
- The involvement of pupils in self-assessment (36%) and peer-assessment (8%). Some reported asking pupils to assess their own performance against stated learning objectives, such as by highlighting ‘I can’ statements, learning ladders, Assessing Pupil Progress grids or level checklists.
- The use of marking to move pupils’ learning forward by explaining their next step, asking challenging questions or identifying ‘two stars and a wish’ where two features are celebrated and one provided as a next step.
- The use of formative evidence to identify gaps in learning and then alter teachers’ planning or provide additional tasks for the children.

The summative strategies reported by the 91 schools represented in the PSQM database (figure 2) tended to fall under the categories of pencil-and-paper testing (red/pink), levelling of pupil work (yellow) and the use of tracking grids (purple/blue) to highlight criteria met and set targets.

An interesting feature of summative assessment was the separation of scientific skills and knowledge by 37% of schools in this sample described a separation of assessment methods, for example by using tests for conceptual understanding and tracking grids for procedural understanding. This could be because the Assessing Pupil Progress (APP) grid (DCSF 2010), still used by 36% of the sample, prioritises scientific process skills within its five Assessment Foci (AFs), requiring schools to find other strategies for assessing knowledge acquisition. Although APP is only a partial solution, tends to be cumbersome, is no longer government policy and unfitted to the 2014 National Curriculum, it remains popular for the following reasons:

“Science APP not only allows the head teacher, staff and myself to track pupils’ progress but it has also helped to maintain the high profile of science in our school following its removal from SATs. It also informs planning and is a valuable tool for ensuring effective differentiation in the classroom.”
(extract from subject leader submission)

“The impact of introducing Science APP has been that staff feel more confident assessing science, assessment is consistent across school, and gives a good overview of a child’s learning and progress in science rather than relying on a snapshot ‘test-style’ assessment.”
(extract from subject leader submission)
Whilst the above PSQM data were collected in 2013, the case studies from TAPS project schools represent current practice since they have been visited as recently as March 2014. Given the imminent disappearance of National Curriculum levels and the fact that APP has not been officially supported since May 2010, we might expect to see some shifts in approach during the intervening year, but it would appear that the broad strategies reported above are still strongly-represented. However, we have found differences in emphasis between schools, which have led to the categorisation of approaches represented in the following case studies.

5.1 A focus on formative assessment with trust in teacher summative judgments
Worlebury St. Paul’s CE Primary School, Weston-Super-Mare

Lisa Dadds, acting head teacher and assessment coordinator at Worlebury St. Paul’s, has based her school’s approach on Dylan Wiliam’s ‘embedded formative assessment’ model (www.dylanwiliam.org). In Wiliam’s model, formative assessment is based on five key strategies in learning:

• Clarifying, sharing, and understanding learning intentions and criteria for success.
• Engineering classroom discussions, activities, and tasks that elicit evidence of student achievement.
• Providing feedback that moves learning forward.
• Activating students as learning resources for one another.
• Activating students as owners of their own learning.

The school’s view is that summative record keeping should be minimal, and formative assessment is where they should be focusing their energy. This follows Wiliam’s view when he says:

“Many schools think that collecting data on their students’ progress and putting it all into a spreadsheet will help learning. There is absolutely no evidence that this kind of monitoring has any impact on students’ learning — as my American friends say, “Weighing the pig doesn’t fatten it.”... [ ]... But the most important assessment happens minute-by-minute and day-by-day in every classroom, and that is where an investment of time, and resources will have the greatest impact on student learning”. (Wiliam 2011)

This emphasis on formative assessment means the school has evidence for many of the boxes in the two base levels of our analytical framework pyramid (pupils’ and teachers’ roles in on-going formative assessment, see appendix). Pupils are involved in discussing learning goals through the collaborative process of constructing a ‘Learning Wall’ as a whole class. A ‘Learning Wall’ is a display board in the classroom that is used to document the development of a topic for the whole class, using children’s drawing and writing and photographs, annotated by the teachers for younger children. (Learning Walls were observed by TAPS researchers in all classrooms in November 2013 and new Learning Walls had been established in the classrooms visited in March and May 2014.) Individuals or groups develop KWL grids (What do I know? What do I want to know? What have I learned?) or Mind Maps that identify relevant prior knowledge the pupils have and what questions they have about the topic. The teacher’s role in this is to bear in mind the expected standards as set out in the curriculum and focus attention on these elements if needed. Science coordinator and Foundation Stage teacher Kate Porter explained that Learning Walls – developed in North Somerset by adviser Tim Sully – have also been an important influence on practice across the whole school.

“We’re really good at ... going from children’s kind of own ideas really ... everyone does a KWL grid [figure 3] at the beginning of every topic. Some people are doing it in science now, so they [ascertain] what the children ... already know, what they need to know and any misconceptions, and that then gives us ... a way forward to plan and create a creative curriculum, which I know that we do really well ... [and] it’s their interest as much as their ideas...”

(science coordinator, Nov 2013)
This sharing of learning goals and expectation was also visible within lessons that were observed, with a range of devices being used to engage the children in discussing or considering these critical/key skills: key questions (What do we need to be like? What will I see? What will I hear?); objective setting (WALT - What we Are Learning Today); use of exemplification (WAGOLL - What a Good One Looks Like); collaboration (Think - pair - share).

Teachers adapted the pace and challenge of lessons in response to the children. In a lesson on electric circuits with Year 1, the teacher challenged a group who had completed the initial task of making a buzzer sound. The teacher had listened carefully to a child’s suggestion that they use a less powerful battery rather than ignoring it to focus on her own aim that they make a switch.

“...Charlie put his hand up and said, ‘well, we need... a battery that's not as powerful, because obviously then the buzzer won't make as much noise’; his ... level of understanding and thinking was really quite super and what we really liked is that you supported his learning. Just because it wasn’t what you wanted, [or] what was in your head, you allowed [him, saying] ... ‘that’s a really good idea Charlie, but it’s still going to be going on’, and he said, ‘oh yes, well you need to switch it off’, and then it developed into the switch.’

(discussion with teacher after a lesson)

Teachers gather evidence of children’s ideas using a range of strategies: teaching assistant (TA) and teacher writing their observations of children’s actions or children’s utterances, TA photographing.

One teacher explained what she had noted:

“I don’t know if you heard Luke; he was saying, ‘if you just attached it to the paper it would stop there and then... if it goes through the metal it will go all the way through’, so he started talking about conductive…”

(Y1 teacher reflecting on lesson, Nov 2013)

Rather than lengthy ‘write ups’ of practical work, they make a careful choice of what children record in written or diagrammatic form, which helps the teacher make a judgment about their understanding. E.g. drawing of a circuit that was made, a ‘graffiti wall’ of a group’s ideas, a poster plan of a design for an ideal habitat (lessons observed Nov 2013-May 2014).

Sample books from every year group show that across the school, teachers are providing children with feedback through their ‘marking’. Teachers write questions that the children write written responses to. Sometimes this is a simple question that requires an answer; sometimes multiple options are offered for the children to circle the idea they think is the best (see figure 4). In every case the child has given some form of response to the marking.

At the next level of the analytical pyramid (see appendix), in order to monitor pupil progress the science coordinator gathers summative levels for each child three times a year. Kate Porter explained that because of the rigorous approach to formative assessment the teachers are not required to produce any further evidence of children’s learning to support these judgments, they are simply asked to record a ‘best fit’ summative judgement of a level of attainment for each child three times a year based on their knowledge of the child.

“Looking at the planning and book scrutiny, they’re obviously getting formative assessment from [identifying] next steps being identified and [the] children responding to that....I think that's pushing them and extending them enough ... We're in danger of putting too much on teachers, and then actually the fun of science goes....you speak to children and they love science, and actually the teachers love doing science as well, and I don't want that to go. I want them to still be excited about it, because that's what it should be all about.”

(science coordinator interview Nov 2013)
The school is clear about where the teachers’ energy is best spent: on formative assessment and on enjoying the science along with the children. This is in line with the ‘Science Principles’ that the staff generated together at the start of the academic year. A similar process has been conducted with the children (as suggested by Primary Science Quality Mark - see figure 5).

“When I first arrived and I was in Year 1 we were doing APP for science for every child. [When] I took over, I said ...we can’t manage this. What is it giving to our children and us? Very little. We’re not taking away anything from it, so what are we doing it for? Everyone agreed, so we stopped it and then I implemented a new system.”

(science coordinator interview Nov 2013)

Kate uses annual scrutiny of work to monitor consistency of expectations across the school:

“We looked at ... evidence of SC1 and we looked at whether it was linked to the planning ... it came up in the summative scrutiny that there needed to be more evidence of SC1 really.”

(interview Nov 2013)

At the level of reporting to parents, student achievement is discussed in terms of what they can do, not only levels or grades. For most year groups, reports to parents are not based on a level of attainment in science, and attitude is an important focus for these reports. At the top level of the pyramid, a level of achievement for science in the year is provided for each child and recorded on a central database (SIMS) for external reporting.

“We give them ... just a one level by the end of the year so we don’t give them a fine level by the end of the year and that goes on our data tracking system.”

(interview Nov 2013)

As the tri-annual recording of attainment science is currently kept by individual teachers on paper or on their personal computers and then collated by the science coordinator, the school will be exploring the potential of an ‘off the shell’ system (e.g. Classroom Monitor) to hold these records.
Science coordinator Kate Porter is also considering how the reliability of teacher’s judgments can be improved and also demonstrated to an external audience. This may take the form of introducing moderation for science, or focussed assessments of ‘working scientifically’. The introduction of the new curriculum in September 2014 provides a natural opportunity to re-establish shared expectations in this new context.

### A variety of Assessment for Learning (AFL) strategies in use in the classroom

**St. Paul’s RC Primary School, Yate, South Gloucestershire**

The science coordinator at St. Paul’s Primary School has been encouraging staff to use formative assessment (or AFL) strategies in science. There was some staff training in 2012-13 on forms of assessment for learning (interview May 14) although the coordinator suggested that to achieve whole school implementation this may need to be revisited as there have been staff changes.

However, a variety of AFL strategies can be seen in use on a day-to-day basis at the school, as evidenced by a number of observations of teaching conducted during 2013-14 in Y1, 2, 3, 4 and 5 across a variety of science topics – bones, rocks, electrical circuits and changes of state.

At St. Pauls, science lessons often begin with a recap of learning from the previous session. Teachers then involve students in discussing learning goals and the standards to be expected in their work (see appendix). This might take the form of a discussion of learning objectives. At this point teachers take care to ensure the children understand the meaning of keywords that will be used during the lesson. So that the pupils can begin to take ownership of the lesson objectives, the children are given an opportunity to discuss them with each other.

A variety of approaches, including interactive whiteboard activities are used as a whole class tool to promote thinking before hands-on practical activity gets underway. Children are targeted to take the lead with some ‘revision’ of previous learning because the teacher has noted they were unsure of a concept last lesson. “Let’s remind ourselves what a ‘fair test’ is ... What do factors or variables mean?”

Once the lessons are underway teachers gather evidence of their students’ learning through further questioning/discussion (see appendix) by using a range of strategies. This might be in the form of partner ‘buzz-time’ discussions, to respond to searching questions such as “What do batteries have inside them? What do you notice (about the batteries)?” Questions are also with individuals or groups as practical work is undertaken.

Teachers will note where the children need to be reminded to focus on learning objectives, and intervene appropriately: “It’s important to explain ...why?” ... Let’s predict what is going to happen ... What are you going to measure?” Opportunities for dialogue might be planned throughout the lesson. Children are happy to seek help, e.g. the child says ‘I’m getting confused’, so the teacher explains in a different way or clarifies the task. Questions are used to encourage children to recap on learning, e.g.: “How did you do that?”

Teachers gather evidence of their students’ learning through observation (see appendix) by planning to work with groups to assess progress or making use of teaching assistants to make observations on specific children as they monitor the remainder of the class. The teacher might say “I’m going to eavesdrop on your group,” as she listens in, and make a post-it note of a key utterance to be used later to assess an individual’s learning. They might draw attention to a child who is making good progress in order to support others to do likewise. “Max – you said something really interesting then – say it again” (recaps learning). One of the adults might take photos of children as they carry out the practical tasks. APP foci are planned (figure 7); for example, evidence for a focus on working methods/working together is gathered through group photos of children working together.

![Image of a worksheet titled “App Assessment Opportunities”](image-url)

---

**FIGURE 7 - MEDIUM TERM PLANNING INCLUDES APP ASSESSMENT OPPORTUNITIES**

Approaches to Science Assessment in English Primary Schools
Teachers gather evidence of students’ learning through study of products relevant to the learning goals (see appendix) during or at the end of each lesson. There might be a ‘mini-plenary’ where a recap of the first activity occurs. ‘Goldsworthy and Feasey’ investigation planning sheets allow for a quick assessment of where each group has reached (figure 8).

**FIGURE 8: PLANNING AN ENQUIRY ACTIVITY USING STICKY NOTES**

Teachers will check progress by reviewing children’s writing, drawing, diagrams and charts and give written feedback. They use assessment to advance students’ learning by adapting the pace, challenge and content of activities (see appendix), perhaps by drawing attention to a child who is making good progress in order to support others to do likewise. If a term is not understood, children are asked to dwell on its meaning. ‘Let’s think about what ‘durable’ means’. ‘Do we mean melt, or dissolve?’

Teachers use assessment to advance students’ learning by giving feedback about how to improve (see appendix) by asking children to apply knowledge learned during plenary sessions (e.g. ‘What is missing from my circuit?’) Teachers note children that have found the concepts difficult and will respond through marking and annotating books (figure 9), revisiting concepts at the next opportunity or amending planning.

**FIGURE 9: AN ANNOTATED AND REVISED CONCEPT CARTOON**

Teachers give feedback during the lessons in a variety of ways (see appendix); in addition to those already discussed they also do this through praise, celebrating children’s small steps and successes. Children are asked to choose a learning objective and say ‘I can…’ Teachers also use assessment to advance students’ learning by providing time for students to reflect on and assess their own work (see appendix) by reviewing learning outcomes during plenaries. A TA might offer feedback on a group’s teamwork skills or an individual’s progress against
the learning objective. Children are sometimes asked 'how well did you do today?' The whole class might look at photos of the lesson in progress and review their learning or return to their KWL grids (figure 10).

The school's next steps are to progress the implementation of AfL by exploring ways in which children can be given a more central role in gathering evidence of learning to aid reflection, inform assessment and facilitate better feedback. To this end St. Paul's staff will trial the 'Live Assess' software made available as part of the TAPS project, making use of a class set of 'FizzBook' computers on loan from Bath Spa University. This will enable children to record key moments in their science learning in a multi-media way, which can then be reviewed by teachers, pupils and their peers.

Strong Assessing Pupil Progress (APP) focus with linked evidence

Drove Primary School, Swindon

Clare Sowden, who came to Drove Primary School as science coordinator in September 2013, brought with her from her previous school an approach to science assessment based upon rigorous use of the APP progression grids (2010) for all children. Whilst Drove Primary staff had been used to English and Mathematics APP grids, Clare's approach for science was new to them. Although she recognised that the 2014 National Curriculum reforms were likely to render level-based systems such as APP obsolete, she felt that setting up an evidence-based system would help prepare staff for national accountability based on teacher assessment to which a high degree of confidence could be attached. This was reflected in the science section of the School Development Plan 2013-14, where Clare acknowledged the need to address ‘... inconsistencies in judgements made using APP and what constitutes suitable evidence,’ and planned to:

‘Develop methods of assessing science throughout the school to ensure judgements are valid and reliable... Moderation of assessments and evidence will confirm reliability of these judgements."

Interviewed in October 2013, Clare identified the two main strengths of science assessment at Drove Primary: first, that staff make notes so that all judgements are evidenced, and second that they are very good at formative feedback. She emphasised that the APP approach can have both formative and summative purposes:

“(You need to) have APP progression sequences ‘in your head’ as you’re feeding back to children, so you know what to go on to next.”

(science coordinator, Oct 2013)

At the base level of our analytical framework (see appendix), Drove Primary makes use of peer assessment to evaluate each other's learning against known criteria and inform next steps. The following examples are drawn from annotations on each other's work:

'I like your explanation about dissolving and your sentence about the liquid dissolving the powder. Improve: maybe label the other two jugs in the picture because I don’t know what's in them.'

(Peer feedback, Nov 2013)

'I like K's technical vocabulary like pollination. The labels are clear and go in the correct order.'

(Peer feedback, Nov 2013)

In relation to the next layer (teachers' role in ongoing formative assessment), Drove Primary’s assessment coordinator identifies a number of strategies in common use across the school:
“So you’ll see speech bubbles in the books…. we also do written feedback… it might just be a question for them to answer, so they’re sort of reflecting on it… make notes on sticky notes of observations of what the children have said, and then there’ll be photographs to go in and support it so that… when we come back to assessment, it’s really clear.”
(assessment coordinator, Jan 2014)

One type of annotation on photos of children doing science is a record of what a particular child said in response to the teacher’s question, which provides important evidence for an assessment judgement, as in the following example:

(T): ‘Why did you choose these objects?
(P): Because they’re all flexible.
(T): How do you know they’re flexible?
(P): Because they all bend.’
(annotation on photograph of Y2 pupil’s sorting of materials)

Another purpose for annotating examples of children’s work is to enable Drove Primary teachers to provide formative feedback to support progression, as below:

‘Extension: check your predictions and mark if they are correct. Next step: to observe what changes happen when a material is heated.’
(teachers’ annotation on child’s grid of predictions about how materials will behave when stretched, twisted, bent or squashed)

In order to track pupils’ progress (the third layer in our analytical frame), Clare has introduced APP grids to colleagues for them to tick off for individual children on a termly basis. Whilst supporting the introduction of these grids in a staff meeting, some of the reliability issues were raised:

“I think the area that people are struggling with is then sort of breaking (science APP) down into sub-levels, so I think that’s an area of focus.”
(assessment coordinator Jan 2014)

To support colleagues in providing evidence for their APP judgements, Clare also introduced a series of assessment booklets linked to particular Assessment Foci (AFs). She explained their purpose and use during interview:

“... it’s got in it tasks that relate to the different units in each year group... they tend to very much either be around using their skills to design something new, or carrying out an experiment to find something out or explaining something, so it does test different skills, but then attached to it, you get the APP points that match that unit.’
(science coordinator interview, Nov 2013)

There is evidence that the APP booklets were welcomed by staff and are in regular use:

“Clare’s given us these like assessments that are linked to the APP... Sometimes it is hard to think of an investigation for some topic areas. So just getting that sort of starter and it worked well because we managed to then differentiate... We do it every term, so six times a year”
(assessment coordinator Jan 2014)

Clare is keen to ensure that colleagues’ APP judgements recorded on the Excel spreadsheets are backed up by evidence, so she has encouraged them to file examples of work, annotated photographs and responses to questions in the assessment booklets such as the following with the spreadsheets and linked to particular AF sub-levels:

“In conclusion the fastest way to melt chocolate was to put it in water... when it is aloud to cool it returns to a solid... Metal: hard strong, flexible, smooth. Wood: hard, strong, flexible, rough. Glass: hard, breaks easily, rigid. Paper: smooth, breaks easily, flexible.”
(child’s responses to APP unit linked as evidence to APP grid)

This evidencing process is not yet electronic, raising some concerns over its manageability:

“And obviously working books photocopied and put in behind the APP sheet.”
(science coordinator interview, Nov 2013)

Annotation, however, provides scope for additional contextual information – such as the degree of teacher scaffolding – to be included, providing a degree of flexibility in the matching of evidence to particular sub-levels:

“... we put things on there like whether they were supported so that, when we track back, and do the APP, we know although it was the independent, things like that...”
(assessment coordinator, Jan 2014)

One particularly impressive feature of the science assessment approach at Drove Primary is the extent to which teacher...
judgements are moderated between groups of three colleagues, three times per year. These ‘triangulation’ groups look at planning, science books and assessment judgements (currently APP levels) together, using a sub-sample of children’s work drawn from three points in the attainment range. As part of the White Horse Federation, Drove Primary colleagues also have the opportunity to moderate with colleagues in other primary schools (figure 11), all of which the Assessment Coordinator believes has improved the reliability of teacher assessment:

“But since APP’s come in, I’ve noticed at moderation that it’s more consistent [at moderation between schools]... So that if you do move schools, you know that a 2c at that school is the same as another school.’

(assessment coordinator Jan 2014)

This rigorous approach to moderation also extends to the handover of assessment records between successive class teachers, thus supporting continuity of expectation between one year group and the next:

“...at the end of the year, we have a changeover, where you’ll sit with the teachers for the next year and you’ll share books and...agree, disagree, moderate together, so that you have got an agreement across all of it...”

(assessment coordinator Jan 2014)

Another part of the pupil tracking system is the transfer of assessment information in the form of sub-levels from the APP grids to the school-wide data management software package SIMS:

“...the teachers go through and put through the sub-level that they are and SIMS translates that into points... and then it calculates how much progress they’ve made.”

(science coordinator interview, Nov 2013)

SIMS data are used to construct reports to parents (the fourth layer of our analytical framework). Whilst this happens three times per year, the level of detail is not currently very high in science, consisting mainly of effort grades and common lists of topics covered for all children:

“We complete reports on SIMS which are then printed for parents three times a year. In terms 2 and 4 science is graded on effort (A = exceptional, B = average, C = requires improvement). In term 6 the report includes science levels.’

(assessment coordinator Jan 2014)

“...the teachers go through and put through the sub-level that they are and SIMS translates that into points... and then it calculates how much progress they’ve made.”

(science coordinator interview, Nov 2013)

In relation to the top layer of our framework (whole-school reporting) however, the presence of detailed science attainment data held electronically on a database such as SIMS enables key staff to manipulate and interrogate these data to monitor progression rates for different groups of children, particularly in a school with a high pupil turnover:

“What I’m doing at the moment is I’m purely looking at Level 3, so I’m tracking all the children that have got a Level 3 at the end of Key Stage 1 and tracking them throughout their year group... And just making sure that the data on there is accurate as well so when someone else flicks through you might notice: hang on, if they were that that term, they’ve gone back there ... it’s just checking with the staff is that right.”

(assessment coordinator, Jan 2014)

“We take that data and do a lot of work on it, track an incidence group for children from Y1 to Y6 to demonstrate that children are making above average progress, because lots come in during the school. Summarise data for governors, account for pupil premium.”

(assessment coordinator, Jan 2014)

Extensive statistical analysis of assessment data held in numerical form needs to be undertaken with caution, since the apparently fine-grained nature of such data is only as reliable as the original teacher judgement which underpins it.

However, Drove Primary’s painstaking approach to evidencing and moderating such judgements provides a level of reassurance on this point. Overall, the National Curriculum levels the school is currently required to submit to central
CASE STUDIES OF TYPICAL APPROACHES TO SCIENCE ASSESSMENT

government are underpinned by high-quality teacher assessment. Even when the national reporting system changes, with levelling and APP rendered redundant, the strong assessment culture within the school is likely to support a smooth transition:

‘I think just people are very good at assessment here. Coming from another school, I’ve noticed that people are a lot more accurate...[a] deeper understanding of what assessments need to be done and what there’s no need to do.’
(assessment coordinator Jan 2014).

Judgements on attainment made using a range of evidence
Brookside Primary School, Street, Somerset

Teachers at Brookside Primary School make judgements of pupils’ attainment in science using a range of evidence collected over a long period of time. This is in contrast to the approach taken by schools that only use end of unit summative tasks to make judgements. The tone for assessment of science at the school is set by the science policy. It states that:

‘Teachers will assess children’s work in science by making informal judgements during lessons. On completion of a piece of work, the teacher assesses it and uses this assessment to plan for future teaching. Written, verbal or visual feedback is given to the child to help guide his/her progress, while older children are encouraged to make judgements about how they could improve their own work... At the end of a unit of work s/he makes a summary judgement about the work of each pupil in relation to the National Curriculum and/or P scales.’
(Science policy)

It is clear that the means by which children record their science work plays its part in enabling assessment to be varied in its nature. At Brookside, evidence of achievement is collected in each child’s topic-work book over the course of a unit of work in science. An advantage of this approach is that pupils can be fully involved in the whole of the assessment process. For example, at the end of lessons, children self-assess their work by commenting on their understanding of both procedures and concepts. Older children record this information in writing, younger children vocalise their thoughts for the teacher to record for them. The science coordinator at Brookside confirmed there were a variety of opportunities provided for children to peer and self-assess work:

“Topic work is peer marked, and the teacher would discuss with children what they should be looking for in the work that they assess. At the end of lessons, children self-assess both their understanding of procedures and knowledge. This happens less when science is taught discretely, more so when science is linked to topic work. Self-assessment in KS1 is more informal but does happen.”
(science coordinator, Nov 2013)

In addition the assessment coordinator explained that:

“Topic books gather a rich range of information for teachers to use in making judgements. Self-assessment is part of the process and is recorded in the topic books – in KS2 children score themselves out of 5 for various criteria and comment on their score.”
(assessment coordinator, Jan 2014)

Opportunities for pupils to self-assess their work are noted in lesson observations made at the school. The self-assessment information gathered can be used by teachers to inform their judgements of pupil progress. The process of recording the comments pupils make about their work varies depending on the age of the pupil. The science coordinator explained that:

“Lower down the school (KS1) teachers make and record comments and observations principally on post-it notes. For older children (KS2) some note-taking of comments also occurs although, principally, assessment is made through marking of work.”
(science coordinator, Nov 2013)

Brookside’s approach of gathering a range of evidence to inform judgements also includes paying heed to pupils’ responses to feedback. Pupils are asked to review and comment on formative feedback comments made by their teacher and the comments made by the pupil can be taken into account in making judgements. Again, the way this is managed is modified to suit the needs of different age groups. The assessment coordinator explained that feedback to KS1 children is given immediately, whereas with older children time is given for pupils to respond to comments made on their work during science lessons.

A typical example of a comment written by the teacher on an older pupil’s work about the concept of water resistance was ‘[You are] beginning to think why they [plasticine shapes]
will be quicker or slower. Can you extend this further?” The sort of verbal feedback typically provided by teachers of younger children in a lesson on properties of materials in Year 1 focused on the teacher listening to children and sharing significant comments made by the children with the whole class. For example, in an observed lesson, the teacher chose to showcase a comment made by one pupil about what makes a material suitable for a particular purpose.

The thoughtful and considered understanding of the relationship between formative and summative assessment that the school exhibits was reinforced by the science coordinator who explained that:

“The assessments are made after each lesson, with summative judgements made at the end of a unit of work using these day-to-day assessments. Judgements are not just based around an end of unit task but, instead, are informed by the formative assessment information gathered.”

(science coordinator, Nov 2013)

From the range of information gathered, scientific knowledge and enquiry skills are assessed against statements on a tracker grid that is included in the pupils’ exercise books. The approach demonstrates how the pupil can be fully involved in the assessment process to the extent that s/he is aware of the criteria used in making judgements.

The tracker system for procedural understanding is divided into National Curriculum levels 1-6 with a series of statements at each level divided into four categories: planning, observation, analysis and evaluation. The statements are expressed in the first person and in a language that makes sense to primary-aged children. An example of a level 1 statement for analysis is ‘I can talk about what has been done’, and a level 4 example is ‘I predict outcomes and compare with actual results’.

The school recognises that not all procedural understanding can be recorded meaningfully or easily in written form in topic books. Teacher observation of procedural skills is required. The science coordinator explained that she wants to see observation playing a more prominent role in assessment, this being particularly the case for older children. Partly as a response to this, during the next phase of the project the school will be trialling an ICT software system that captures learning of practical elements of science in a variety of modes (video, still image, audio etc), then reviewing the system’s effectiveness in supporting teachers’ and learners’ assessment of enquiry skills. As the ICT coordinator explained:

“Throughout the school, [pupil and teacher] comments and evidence have been gathered separately and in time consuming way so the hope is 2Simple [the ICT software system the school will be trialling] will make this whole process more manageable, accurate and instant. It seems pupil-friendly so, further up the school, the system could be used by pupils to record experiments. It will probably be used mostly as a formative assessment tool. Hopefully, employing it throughout the school will improve consistency. Potentially it could be used in many different areas of learning, not just science.”

(ICC coordinator interview, March 2014)

During a lesson observed in a Year 2 class the teacher provided an opportunity for pupils to reflect on and assess their own work during a ‘mini-plenary’. This gave each child a chance to reflect on the way they had set up their investigation. During the plenary, children evaluated the quality of their scientific procedure. In a lesson observed in Year 6, pupils were continually making informal judgements about their work. It was easy for children to judge their success in the lesson, which was focussed on separation of materials. Pupils simply observed the extent to which the different materials had been separated. In addition, groups compared work with one another.

Assessment at Shaw is based on a shared understanding of “...what good science looks like” (Subject leader interview, Nov 2013). This is the case at all levels of the school, with pupils and class teachers very clear about what they should be aiming for. Science subject leader (and Fellow of the Primary Science Teacher College) Carol Sampye ensures that enquiry is key and each class displays science ‘star’ guidance and a skills wheel (designed by Wiltshire Science Adviser, Tom Robson). The skills wheel supports coverage of enquiry skills so that it is clear to the children which area is the focus for that lesson or sequence. The science ‘stars’ provide details of, for example, what a 4* scientist looks like - the key features of a level 4 in scientific enquiry. The science stars are also embedded in the planning, which details success criteria for the sequence of lessons. The strong enquiry structure supports on-going
Formative assessment to move the children forward each lesson, but also enables the staff to be confident about making summative assessments. Carol is keen that levels are not based on testing:

“There is agreement that assessment should not be rigid or an exercise in filling in boxes. Paper testing is limiting and does not necessarily give an accurate measure of attainment.”

(science coordinator interview, Nov 2013)

Teachers use evidence gathered in pupil books or group floorbooks, together with their knowledge of how the children have responded in lessons, to provide a best-fit level at the end of each topic.

Staff confidence in levelling is supported by regular moderation. Carol set up a series of 10-minute science moderation slots, which take place within staff meetings across the year. Each slot consists of one class teacher bringing along some samples of work, which could be children’s writing, drawings or speech, and the staff agreeing a level for each piece. This led to lengthy discussions at first, but the process became quicker as staff developed knowledge of what to look for:

“Moderating regularly in small manageable chunks helps us to maintain a high profile for science, gives teachers confidence and means we have super evidence of children’s attainment.”

(subject leader, PSTT exemplar)

Such moderation has led to the creation of a portfolio of evidence, examples of which can be found in a PSTT exemplar at: http://www.pstt.org.uk/science-teaching/primary-science-quality-mark/exemplar-materials.aspx

A high emphasis is put on speaking, listening and group work which is evident throughout the school. This is combined with a focus on self and peer assessment in many lessons. For example, in Year 4 the children had found it difficult to create a branching key in their books, so the next lesson began with pupils making keys with ‘sticky notes’ in small groups. After talk partners had formulated yes/no questions to divide the animals, the children identified for themselves if they felt confident or not, with those who found it tricky being placed with those who did not. After ten minutes creating their own keys, the pupils were given time to walk around to look at others’ work. Rather than the teacher picking out good ideas, these were put forward by the pupils who then went back to improve their key.

In a Year 5 lesson on Space the children were physically modelling the orbit of the Earth and the Moon using different-sized balls. As they moved the balls they gave a commentary about what was happening, which was then peer assessed for clarity and accuracy. The groups gave advice to each other on how to improve their explanation to ‘become 5* scientists who use scientific vocabulary accurately’. This exemplifies assessment at Shaw: a clear framework for progression which is shared and used by both staff and pupils, enabling assessment to be used FOR learning.
Although it would be dangerous to generalise from our sample of 91 primary schools submitting to the PSQM, plus the 12 TAPS project schools, we have been able to form a picture of primary science teacher assessment across England through supplementing these data as follows:

- Visits to four additional schools where science is being led by Fellows and Members of the PSTT Primary Science Teacher College (Stockport, Lancashire, Lincolnshire, East London).
- Responses to presentations of our findings at the Primary Science Teacher College annual conference, teacher conferences in Oxford and Newbury, and the annual conference of the Association of Tutors in Science Education (ATSE) and National Advisers and Inspectors Group for Science (NAIGS) during spring/summer 2014.

Overall, our findings suggest that, in England:

- There is a wide variety of practice in primary science teacher assessment. This diversity has been encouraged by the UK government, firstly by the removal of science SATS at Key Stages 1 and 2 and more recently by the lack of prescription on assessment in the 2014 National Curriculum. Schools have been invited to apply for grants from an Assessment Innovation Fund (DfE 2013) and to share their practices more widely through the DfE website.

- This encouragement of diversity appears to be promoting a range of creative approaches to formative assessment of children's scientific skills and knowledge, from ‘Learning Walls’ to effective teacher feedback and peer assessment. It is unclear, however, how this rich array of formative assessment strategies is feeding back into teacher planning, and there is very little evidence of formative assessment being used to inform summative judgements, as recommended by Nuffield (2012).

- The lack of centralised prescription appears to be having a detrimental effect on summative assessment and the tracking of pupil progress. Many schools remain wedded to bureaucratic numerical systems - either electronic or paper-based - which record children’s National Curriculum levels, sub-levels and ‘average point score’ (APS), breaking down the original levels 1 to 6 into minute sub-divisions with little evidence to support the validity or reliability of these judgements. Having bought into a previously government-sponsored scheme - such as APP or RaiseOnline - or a commercial tracking system such as Rising Stars, SIMS or Classroom Monitor, most schools appear very reluctant to let them go. This could be from fear that the lack of numerical tracking will expose them to criticism from Ofsted or local authorities in relation to accurate ‘measurement’ of pupil progress, or that they can’t quite believe that the levels really are disappearing.
The five snapshots of school practice in science assessment provided by the case studies above, whilst all differing in the tools used and the ways in which children's progress is tracked, display some common features which our analytical framework (see appendix) would suggest exemplify good practice:

- **A strong emphasis upon formative assessment** (AFL) as lying at the heart of the teacher assessment process and which leads or drives the summative judgements made. The use of Learning Walls, KWL grids, ‘buzz’ groups, exemplification of objectives and IWB discussions all have high validity as assessment strategies, though recording them more formally raises manageability issues for teachers.

- **A concern to involve children as much as possible** in assessing their own science progress, providing feedback to each other and responding to the interactive feedback of their teachers and TAs. This is consistent with the messages from Hattie’s (2008) meta-analysis of educational research that feedback acted upon has the largest effect size on educational outcomes of any intervention. Teachers’ comments are likely to have greater reliability (especially if moderated) than peers’, but both can have positive impact on children’s learning.

- **A separation between the assessment of procedural and conceptual components** of scientific attainment, which was also reflected in the PSQM data. This increases the manageability of the assessment process, but arguably compromises its validity, as scientific process skills may be concept-dependent so need to be assessed in relation to a range of conceptual content.

- **A rigorous approach to evidencing teacher judgements** – particularly in relation to the use of Assessing Pupil Progress (APP). Unlike the high reported use of APP spreadsheets by PSQM schools, only one in five of our case studies made extensive use of this scheme, suggesting either recognition that it is no longer government policy or concerns over its manageability. Certainly, evidencing every judgement with a piece of children’s work, an observation or quote would create an unmanageable system, but a light sample of evidence can provide assurance of the consistency (reliability) of teachers’ judgements and the validity of assessment activities.

- **A focus upon moderation of teacher judgements** as part of the transfer of evidence gained from formative assessment to quantitative tracking systems, thus increasing the reliability of those judgements.

There is much to be applauded and drawn from these schools’ approaches to assessment of children’s scientific learning, not least the commitment to staff development to enable all colleagues - teachers and TAs - to gain a good ‘feel’ for what it means to be a scientist. At present, some aspects of this ‘feel’ have been for a ‘levelness’ in National Curriculum (1999) terms, so there is a job to be done to translate that into 2014 age-related expectations. However, the sound principles of assessment being laid down in these schools should enable this translation to occur relatively smoothly.
To support schools in making the links between formative and summative assessment, and to move from numerical tracking systems to more valid and reliable strategies for tracking pupil progress against national age-related statements of attainment, the TAPS project will, during 2014-15:

- Exemplify every cell within our model of ‘best practice in science teacher assessment’ (see appendix) with at least three examples of good practice drawn from TAPS schools and others we visit. The model, with hyperlinks to its bank of exemplification, will be made available via the PSTT website for schools to use.
- Develop the above model into a school self-evaluation tool, available online or as an ‘app’ to support schools in auditing their own assessment practice and identifying gaps or areas where reliability or validity could be improved.
- Begin to develop a ‘standards file’ of pupil work against each of the age-related statements of attainment in the 2014 National Curriculum for Science (Years 1 to 6), using our project schools to generate and moderate examples of ‘emerging, expected and exceeding’ statements either individually or in clusters. This will be made available online.

REFERENCES


APPENDIX: Analytical frame for case-study data
(and whole-school science assessment self-evaluation tool)

Produced by the Teacher Assessment in Primary Science Project,
Bath Spa University, developed from the
Nuffield Foundation (2012) and Harlen (2013)