Davies, D., Collier C., Earle, S., Howe, A. and McMahon, K. (2015) Preparing Science Teachers for Assessment Without Levels, *Science Teacher Education* 72: 9-16.

Introduction

Assessment is primarily a matter of judgement rather than measurement, yet for too long we have been pretending that we can measure pupils' attainment and progress in increasingly fine detail (one APS 'point' being one sixth of an original National Curriculum level). The lack of validity and reliability of this approach becomes obvious when we try to assess something as multi-dimensional as practical work in science (Roberts and Gott 2006), yet the current 'bonfire of the levels' in the new National Curriculum in England (DfE 2013) has left schools and teachers feeling vulnerable and reluctant to discard the 'comfort blanket' of numerical tracking systems. Data from our Teacher Assessment in Primary Science (TAPS) project suggest that very few primary schools have yet adapted their assessment approaches to the 'post-levels' world, and that most will continue levelling pupils during 2014-15 – as indeed they are required to for Years 2 and 6 - whilst possible alternatives are explored. We suspect that the situation is similar in most secondary schools, particularly as they prepare for the introduction of the *Progress 8 school performance measure* (DfE 2014), which aims to track the progress of pupils from the end of KS2 to GCSE. On our primary and secondary PGCE programmes at Bath Spa University we have in the past introduced beginning teachers to a number of formative strategies for science assessment, yet saved the summative process of 'levelling' to the end of the course, since this is one of the areas they find most difficult.

So we should see the loss of levels as an opportunity rather than a threat; to bring formative and summative assessment closer together and ultimately to find more valid ways of assessing what it means to be a scientist. The TAPS project, based at Bath Spa University and funded by the Primary Science Teaching Trust, aims to develop a system for assessing science which will support teachers to use the full range of pupil information available in the primary classroom to assess and develop learning. The research questions we are seeking to address are:

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?

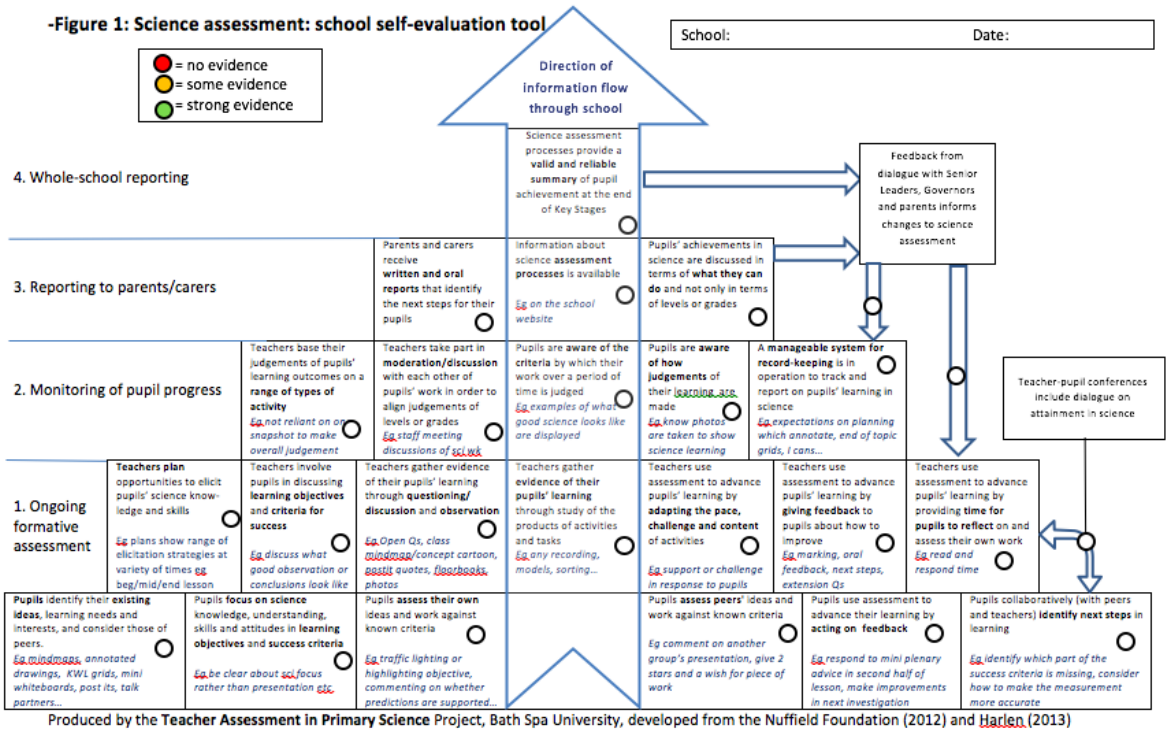
Our findings to date are based on analysis of two principal data sources:

1. The submissions to an online database of science subject leaders in all 91 English primary schools who worked towards the Primary Science Quality Mark (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 co-ordinators, 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.

The model of teacher assessment developed through the TAPS project

A working group of science assessment experts convened by the Nuffield Foundation (2012) recommended that the rich formative assessment data collected by teachers in the course of ongoing classroom work in science should also be made to serve summative purposes (reporting to parents, teachers of the following age group, government) through synopsis at the end of academic years or key stages. They developed a pyramid model for the flow of assessment information through a school, using the analogy of energy flow through a pyramid of numbers in an ecosystem. The TAPS project aims to operationalize the Nuffield working group recommendations by developing this pyramid model into a whole-school evaluation tool (Figure 1, available on www.pstt.org.uk) to support schools in identifying strengths and weaknesses in their assessment systems and provide an exemplified model of good practice. Assessment information feeds up from the ongoing formative assessment layers with the actions of pupils and teachers in the classroom being the basis of later monitoring or reporting. In the same way as feedback loops within an ecosystem affect populations in the layers below, feedback from summative assessment, tracking and reporting can influence how teachers and pupils make use of formative evidence.



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

To exemplify the layers and cells within the pyramid tool, we have gathered examples from project schools, published as a series of case studies in Davies *et al.* (2014). For example, in one school, children are involved in discussing learning goals through the collaborative process of constructing a 'Learning Wall' as a whole class (see base layer of Figure 1). Individuals or groups develop KWL grids (What do I *K*now? What do I *W*ant to know? What have I *L*earnt?) or Mind maps that identify relevant prior knowledge the children have and what questions they have about the topic. 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.

In another school, *teachers involve children in discussing learning goals* and the standards to be expected in their work (see second layer of Figure 1). At this point teachers take care to ensure the children understand the meaning of key words that will be used during the lesson, giving them an opportunity to discuss them with each other. Once the lessons are underway *teachers gather evidence of the children's learning through further questioning/ discussion* 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)? 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. *Teachers gather evidence of the children's learning through observation* 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.
In relation to the third layer of Figure 1 (monitoring progress) another school's approach to *gathering a range of evidence to inform judgements* includes paying heed to children's responses to feedback. 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. From the range of information gathered, scientific knowledge and enquiry skills are assessed against statements on a tracker grid that is included in children's exercise books. The approach demonstrates how the child can be fully involved in the assessment process to the extent that s/he is aware of the criteria used in making judgements. The statements are expressed in the first person and in a language that makes sense to primary-aged children. In another school the science subject leader set up a series of 10-minute science moderation slots, which take place within staff meetings across the year: '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). This moderation has led to the creation of a school portfolio of assessed work in science.

At the level of reporting to parents (level 4 in Figure 1), in one school children's achievement is discussed in terms of what they can do, not only in terms of 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. In relation to the top layer of our pyramid (whole-school reporting – Figure 1) the presence in another school 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 turnover. 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 judgements which underpin it, however the school's painstaking approach to evidencing and moderating such judgements provides a level of reassurance on this point.

Overall, whilst differing in the tools used and the ways in which children's progress is tracked, science assessment in the schools we have visited display some common features which our evaluation tool 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.
- A separation between the assessment of procedural and conceptual components of scientific attainment. 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. Clearly, evidencing every judgement with a piece of children's work, an observation or quote can 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, particularly if hyperlinked using an electronic system.
- 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.

One final feature of these school's approaches to assessment of children's scientific learning, is 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 'levelness', so there is a job to be done to relate this to 2014 age-related expectations or performance descriptors.

Implications for Initial Teacher Education

Although we have not reached the stage of the PGCE programmes at Bath Spa University where we introduce our beginning teachers to the mysteries of summative assessment, we are conscious that this needs to be done in quite a different way from previous years in order to avoid inducing in them the fear and confusion currently affecting many schools. By introducing them to the principles of good practice in teacher education through the pyramid model; exemplifying each layer and cell for them and inviting them to use it as an evaluation tool in their own classroom practice we aim to develop a new generation of teachers who are confident in their exercise of assessment judgements and are not reliant on computer-based, quantitative level-tracking software to validate their professionalism.

References

Davies, D., Collier, C., Earle, S., Howe, H. and McMahon, K. (2014) *Approaches to Science Assessment in English Primary Schools,* Bristol: Primary Science Learning Trust. [Available online: http://www.pstt.org.uk/Files/Project_Files/ASAEPS%20full%20FINAL1.pdf, accessed 27.11.14]

Department for Education (2013) *Science - Programmes of study for Key Stages 1-2*. London: DfE.

Department for Education (2014) *Progress 8 school performance measure: Information for school leaders and governing bodies of maintained schools, academies and free schools*, London: DfE.

Harrison, C. and Howard, S. (2009) *Inside the Primary Black Box.* London: GL Assessment

Nuffield Foundation (2012) *Developing Policy, Principles and Practice in Primary School Science Assessment,* London: Nuffield Foundation.

Roberts, R. & Gott, R. (2006) Assessment of performance in practical science and pupil attributes, *Assessment in Education: Principles, Policy & Practice*, 13(1): 45-67.