Teacher Assessment in Primary Science (TAPS) Project

Evaluation Report 2015-16
Dan Davies, Cardiff Metropolitan University

Introduction

This brief report is based on interviews with science subject leaders from six TAPS project schools (50% sample) during the final cluster day on 28 June 2016, as part of my role as project evaluator. The questions were designed to elicit the impact that participation in TAPS had made on participant schools' approaches to science teacher assessment – particularly in enhancing validity, reliability and manageability; their developing understanding of the relationship between formative and summative assessment in science; their use of TAPS outputs including the 'pyramid' self-assessment tool and focused assessment tasks; and any noticeable impact on children's learning. These themes are closely related to the original project aims and were translated into an interview schedule which can be found in the appendix to this report.

Sample - *anonymised*

School name	Date of joining TAPS	Interviewee name(s)	Interviewee role(s)
A	Sept 2013	Subject leader A	Science subject leader Science subject leader
B	Sept 2015	Subject leader B	Science co-ordinator and Reception teacher
C	School Sept 2013 SL Sept 2014	Subject leader C	Science co-ordinator Computing co-ord
D	Sept 2013	Subject leader D	Science subject leader KS2 co-ordinator
E	Sept 2013	Subject leader E	Science co-ordinator
F	School Sept 2014 SL Sept 2015	Subject leader F	Science subject leader

Methodology

Interviews were audio-recorded and transcribed. Nvivo qualitative analysis software was used to code the transcripts.

Findings

The 'code-tree' indicating the codes used, their hierarchical relationship, the number of sources (transcripts) in which they appear and the number of references (quotes) linked to each code is in Figure 1. This indicates that the most commonly-used codes were 'impact on children's learning', 'impact on science in school' (both 8 references from 4 sources) 'manageability' (8 references from 5 sources) and 'moderation' (part of 'improving reliability' – 6 references from 5 sources).

Figure 1: Code tree for TAPS interviews (n=6)

Name	Sources	References
⊟ ◯ focused assessment tasks	2	2
◯ Impact on manageability	2	2
◯ Impact on reliability	2	2
◯ Outcomes from tasks	1	1
◯ relationship between tasks and curriculum	1	1
⊟ ◯ Formative	3	4
◯ Comparison with summative	3	3
◯ Relationship with summative	3	4
◯ Understanding of	1	1
⊟ ◯ Impact on children's learning	4	8
◯ Chilcren's understanding of science	1	1
⊟ ◯ Impact on personal pedagogy	1	2
◯ Role within school	1	1
⊟ ◯ Impact on science in school	4	8
◯ Consistent approach to assessment	3	8
◯ Impact beyond school	1	1
◯ Impact on transfer to secondary	1	1
◯ Teacher's understanding of assessment	3	6
◯ Learning from each other	1	1
◯ Manageability	5	8
◯ Profile raising	1	3
⊟ ◯ Pyramid	2	2
◯ Exemplification	0	0
◯ Extension to other curriculum areas	1	1
◯ Use as school self-audit	3	3
◯ Use in CPD	1	2
◯ Use in development planning	2	2
⊟ ◯ Recording	1	1
◯ Multi-modal	3	5
⊟ ◯ Reliability	3	5
⊟ ◯ Moderation	5	6
◯ cross-school moderation	1	1
⊟ ◯ Summative	4	6
◯ Assessment criteria	1	1
◯ Relationship with formative	2	2
⊟ ◯ Tracking systems	4	4
◯ evidence	3	4
⊟ ◯ Validity	2	2
◯ Assessment criteria	1	1
◯ Improving validity	5	5
⊟ ◯ Working scientifically	1	2
◯ Relationship with conceptual assessment	2	3

Impact on children's learning

This was viewed in terms of increased enthusiasm, confidence and ownership:

> I think it's just reinvigorating what science is… igniting the spark back into children, the enjoyment of science

> they've got a bit more ownership for what they're doing as well, and they know where they are, they know, my class loves science and they're confident at their own ability as scientists

> I think they know they're scientists. Absolutely, that's the first thing I'd say. Children in our school believe they are scientists… they know they have those skills and that they are working scientifically.

> It's not happening to them; it's happening with them and by them.

There was also a sense of children being able to provide multi-modal evidence of their scientific attainment:

> It's allowed them to express it in different ways… if we ask them to explain, to show, to demonstrate, to act out, to video-record, any of the more visual, more audio, they are much better at it.

Impact on science in school

The TAPS project was seen as instrumental in raising the profile of science within project schools, by linking the visibility of the subject area to its assessment:

> I think it's had a fresh approach into this school to make science more of a higher priority than it had been.

> Has been a great prompt for us to refocus some of our school improvement on developing science

> we've got a high profile of science in school now, so we're in a good position to be able to start thinking how we assess.

> It's raised the profile of science in school also made us think about how we are assessing it.

The project has helped to share good practice in science learning and teaching across each school, providing opportunities for teachers to work collaboratively:

> And it gave us a forum for conversation. It opened up the professional dialogue. Instead of us hiding in our classrooms and trying to do it on our own, we started swapping ideas with recognition that failure's a good thing

> It became very evaluative and very experimental, and we tried different things.

Unsurprisingly, much of the impact of TAPS has been on improving consistency of approach to assessment in project schools, including consideration of the relationship between conceptual and skills assessment and how children demonstrate their capabilities:

> Developed 'big ideas' approach to science assessment. Written a working scientifically big idea which we hope to use to drive improvements

get our staff to just put assessment for learning into everything they're doing, really, and to work in different ways.

a consistent approach to assessment, which is focused on assessing the ability of the children to work scientifically

map out… each conceptual understanding and match it to a working scientifically focus so that each term, we focused on two as a whole, and that's what we recorded in our books and that's the evidence we collected,

For years there wasn't much recording of science assessment, but now we have a reasonable online system for school.

The impact has been—in order to access all the areas of the assessment, we are ensuring that we are giving opportunities for children to demonstrate theirs. Particularly at the top end of the school, where there's greater independence required, they have to demonstrate that they have, with understanding, selected this system or selected that piece of information or present it in that way.

Improving manageability of assessment

By helping teachers to focus on 'what counts' as evidence of scientific learning in their planning, the project has led to 'slimmed down' approaches to recording, monitoring and tracking assessment that avoid bureaucratic collection of extraneous data:

I think it is manageable. It's about how you plan your activities at the start and are you having these assessment opportunities and do you know what they're for and how they're going to inform other things.

We have had honed it down a little bit… they now identify from each lesson both the skills, working scientifically skills, and the knowledge that is being developed in that session and what the success of the criteria is going to look like

By being so focused in what we're actually trying to teach, they know, therefore, what they're looking for in the learning, and it's more manageable for them because the children know what to demonstrate

We've found more efficient ways of doing that and trusting our judgment, each of those judgments, and saying, "I heard him say that. I can't possibly write it all down," because you can't, "but I knew what I was listening for, and I'm satisfied that that child said and did whatever it was that was required to match that. I've just ticked it.

The provision of assessed examples of children's work – either through the focused assessment tasks or the online exemplification of the pyramid tool – was seen as instrumental in both improving reliability and making the process more manageable:

You can't make it manageable if there is nothing to compare it with, so you need to start with examples of how pupils are learning

I think it will become more manageable with this focus assessment task because then everyone is just doing a specific part… You need something that everyone is working from across the whole school so that it can be reliable and it can be a good use of assessment. They seem like they're going to be straightforward and you've got lots of examples coming online, so it'll be good.

Improving reliability of assessment through moderation

Although many of the schools had used moderation previously, they suggested ways in which it had become more focused across (and between) schools to give greater confidence to teachers' judgements and indicate areas for further focus:

> I guess it's the internal moderation. Then we pick up things, like the working scientifically, if that's an area that we need to develop and work on.

> I think we've done more moderation; we've done some cross-school moderation within our cluster and delivered to them what we've learned as a Taps project, and we've all looked at books of similar (levels of attainment)

> we have done several snapshots where we have staff meetings either brought work together that we thought was shown working at their eye level or we brought work where we divide it and moderate and validate it for ourselves… we've done a lot more within each key stage. We've got pairs of teachers working together who have validated those judgments at the end of units

> I have done a little cross-moderating with my cluster group in that we've brought work together, and I've taken work from our school that we've agreed has been a certain level and said to the cluster group, "Why do you think we think this? do you agree? Why is it different? What might it look like in your school?"

Improving validity of assessment

This code was used 5 times in 5 different interviews, but generally there is surprisingly little evidence of more valid approaches to assessment. There still seems to be some confusion amongst some participants concerning the difference between validity and reliability in assessment. Others appreciated that multi-modal approaches to collecting evidence of scientific learning could improve validity:

> There are different ways of showing the children understand what they're doing. It could be them doing a video presentation on something they've learned about the life cycle of a plant or working collaboratively and recording what they're doing and how they may be solving problems through the investigation as they go along rather than just writing down what you found out.

It was also acknowledged that by being more specific and focused on what skill/concept was being assessed this could improve validity:

> As long as it's linked to the objectives and you're not just going off and assessing everything—it has to be focused and you need to know where the next steps are and what the children can do—then, yeah, absolutely, it's very valid

Impact of the 'pyramid' self-assessment tool

Although the code 'pyramid' was only used twice, the different ways in which schools had made use of this key output of the TAPS project totalled 9 references. Participants had found it useful as general self-assessment of science pedagogy across the school - not just in terms of assessment – and for individual teachers to use to reflect on their practice. It had also supported science development planning:

For me personally, the pyramid is quite a good starting point as a coordinator to see which bits we're doing really well, which bits I need to focus on next with the staff to move the subject on

finding out about how assessment worked in our school… to find that our assessment was lacking slightly, finding out how to improve standards across, and with the implementation of the new curriculum, trying to find out how we can make science more of a priority

The pyramid we used initially and in its various reincarnations to do an evaluation as a school of where we all thought we were at and for the class teachers at the same time to do an evaluation of where they felt they were at.

I shared it with senior leadership of how I want to see science moving forward

Two of the science subject leaders had used the pyramid tool as part of CPD for colleagues:

I've tried to share lots of, the more the bottom layers of the pyramids I've been doing with the school, trying to show them, give them lots of examples whenever we have staff meetings, try and give them lots more assessment for learning techniques…

To be honest, the most important thing with the pyramid is to give the teachers examples to click on and see different ways of doing it

The relationship between formative and summative assessment

Our baseline assessment of the participants in TAPS indicated that they tended to separate the formative and summative purposes of science assessment in their schools without making explicit links between them; a finding echoed by the analysis of submissions to Primary Science Quality Mark. Although there were only four statements from three schools which were coded in this way, there is some evidence of an emerging understanding of the potential for use of evidence collected to support learners being summarised for monitoring, tracking and reporting purposes. This sometimes resulted in a cessation of summative testing:

I guess it's when we use our tracking system; that then becomes your summative assessment. But you've done the work before that. There are lots of different objectives within one unit on plants or one thing on electricity. By looking at each objective as formative assessment, you can then see what their overall understanding of that particular unit is, if that makes sense.

The main change is that our assessment is ongoing. We don't do any summative testing at the end of unit, so at the end of the year, we are continuously gathering data, more information about the children that informs a consensus of an idea at the end in terms of offering our head teacher or our management a summative grade.

… what we're doing now is doing ongoing formative assessments throughout a unit of work, and at the end of each piece of unit of work… and we use those judgments at the end of each unit of work—and that's both the working scientifically and the conceptual knowledge—to inform an annual judgment about that child, which then goes towards a summative statement that is passed on to the next teacher and then used as a summative statement for the end of the key stage, which had been used at the end of key stage for tracking purposes.

There was acknowledgement from one participant that the relationship could work in the opposite direction, with summative assessment potentially being used for formative purposes:

> We feel strongly that even summative assessment has to have a formative purpose. We want teachers to be asking 'how will the child completing this focus assessment task help us to improve teaching and learning for them and for future cohorts?' Making a judgement about the level that individual children are working at is a secondary outcome

One teacher had developed a 'snapshot' view of summative assessment – rather than an accumulation of evidence over time leading to a summary judgement. This participant regarded formative and summative assessment as qualitatively different, yet combining to give a (more valid?) overall picture of a pupil's learning:

> Summative assessment is really only a snapshot. Some children for example if you're doing an end-of-unit test and they're not performing to their ability, using the test score in combination with your formative assessment gives you a much better overview. So they work with each other really

Summary

This brief evaluative 'snapshot' – drawing upon a limited dataset collected during a single morning at the end of the project – suggests some of the ways in which the TAPS project has made a difference to science learning, teaching and assessment in project schools. These data could be combined with the many other sources of evaluative evidence (including participant questionnaires, observations during school visits, school documentation, responses of the hundreds of teachers who have attended presentations and CPD events) to give a comprehensive evaluation of the TAPS project. It is particularly encouraging that the project outcomes are still being used in project schools to drive change in science assessment towards more manageable, reliable and valid approaches which draw upon a developing understanding of the relationship between formative and summative purposes.

Appendix – interview questions

1. What has been your role in the TAPS project?

2. What have been the main benefits for your school of being involved in the TAPS project?

3. What changes to science assessment have you made across the school since the beginning of the project?

4. Please outline the relationship between formative and summative assessment of science in your school

5. Can you give any examples of how science assessment in your school has become more valid?

6. Can you give any examples of how science assessment in your school has become more reliable?

7. Can you give any examples of how science assessment in your school has become more manageable?

8. Can you give any examples of how science assessment in your school has made a positive impact on pupils' learning?