Teachers' attempts to improve assessment practice in primary science are influenced by job role and teaching experience

● Isabel Hopwood-Stephens

Abstract

Recent changes to assessment policy in England have brought the development of primary teachers' assessment literacy in science to the fore. The TAPS pyramid is a tool to help teachers and schools improve their assessment practice in primary science. It has been downloaded thousands of times across 45 countries, but little was known until now about its impact upon the assessment practice of the teachers using it.

This report analyses quantitative data from an online survey of 96 teachers using the TAPS pyramid to show that changes in practice vary across job role and teaching experience. These differences are explored with reference to changes in national assessment policy, but also the wider international research into developing primary teachers' assessment literacy. Finally, an argument is made for school leaders to consider the diversity in assessment literacy present among their teachers when developing primary science assessment practice.

Keywords: Assessment practice, primary science, role, experience, variation.

Formative assessment in primary science

Formative assessment has been defined by Klenowski (2009) as: '*[the] everyday practice by students, teachers and peers that seeks, reflects upon and responds to information from dialogue, demonstration and observation in ways that enhance ongoing learning'* (p.264).

The proven power of formative assessment to improve teaching and learning across the curriculum (Black & Wiliam, 1998a, 2009) has led to a gradual shift in attention from summative written assessment as the way to judge pupil progress, to the ongoing use of formative assessment by teachers to '*identify specific student misconceptions and mistakes while the material is being taught'* (Kahl, 2005, p.11).

In the UK, today's primary school teachers are expected to be skilled practitioners of formative assessment (Ofsted, 2013). Indeed, the importance of developing teachers' assessment literacy during initial teacher training has been both recognised for its importance and lamented for its variability (Carter Review, 2015).

Formative assessment is an intentional form of assessment (Hondrich *et al,* 2016). Unlike a written test with a fixed marking scheme, it is a dynamic process mediated by the teacher, who will plan appropriate opportunities to use strategies such as questioning or elicitation, reflect upon their outcomes and use those to shape further input, both 'on the fly' while teaching (Serret *et al,* 2017) and afterwards, while marking students' work or planning further lessons.

Subject-specific guidance on how to use formative assessment effectively in the teaching of primary science has been available for over a decade (Black & Harrison, 2004), and various formative assessment strategies appropriate to the teaching of primary science have been identified in the literature (Hodgson & Pyle, 2010).

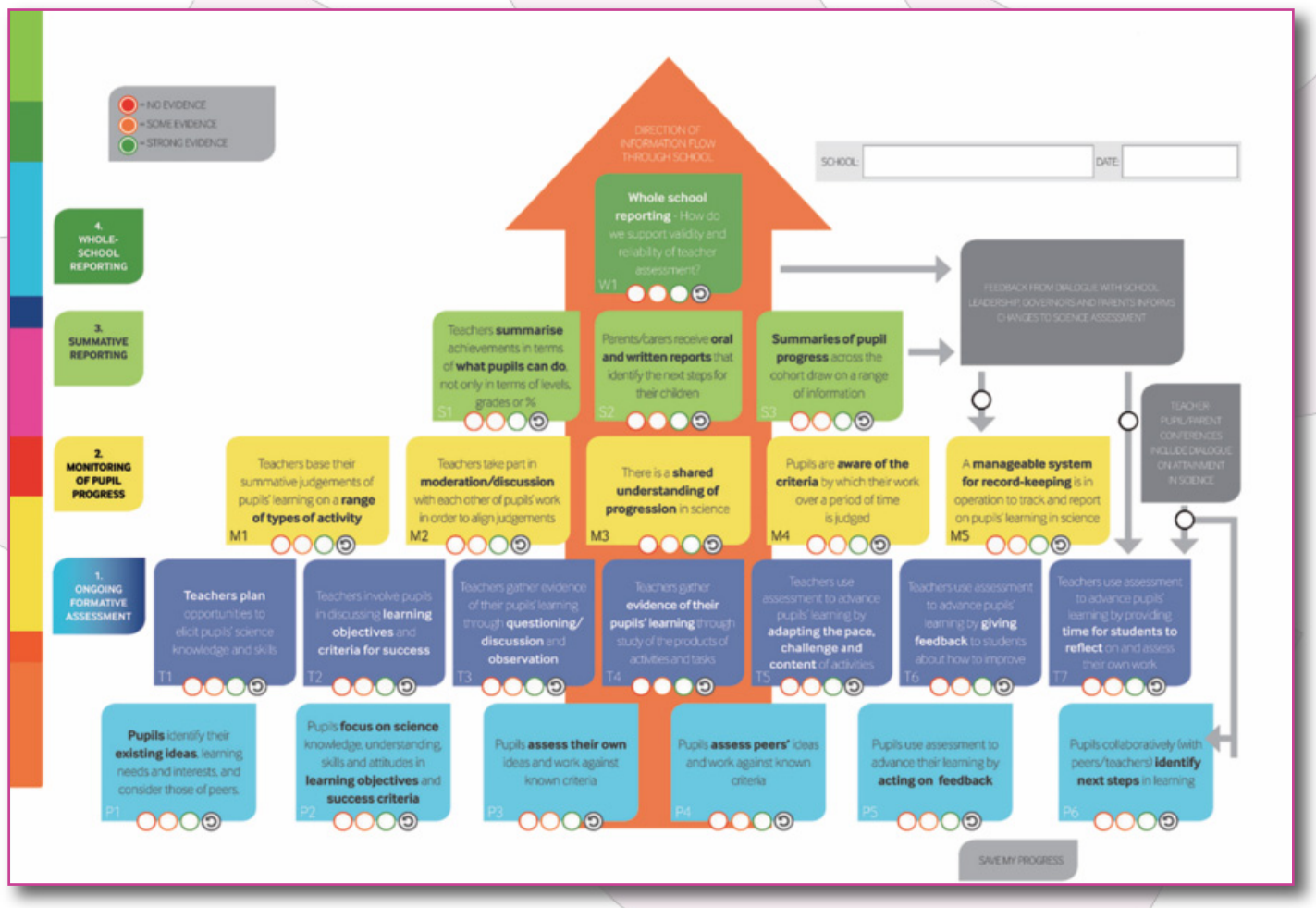
UK primary teachers have been shown to use formative assessment strategies considerably less in primary science than in other core subjects, however (Hodgson, Pyle & Shamsan, 2009). To understand why this issue might have arisen, it is useful to understand recent changes to assessment policy and the curriculum for primary science. Thirty years ago in England, summative judgements of pupil progress were produced by the class teacher, using a range of sources and examples of work. This changed in 1988 when, in

an attempt to standardise the criteria used and the judgements themselves, learning outcomes for primary school pupils in science were assessed through externally administered high-stakes written tests in the final year of primary education. Progress towards these final summative tests was monitored using a system of best-fit descriptors of ability in various skills and knowledge areas, a process known as levelling. The perceived need among school leaders to achieve good outcomes for their pupils in these national high-stakes summative tests, as well as to demonstrate progress against the levelling descriptors, led many schools to rely upon regular written testing of knowledge to demonstrate that learning was taking place (Tymms, Bolden & Merrell, 2008).

Concern about how such testing was distorting the primary curriculum led to the abolition of summative assessment by high-stakes written test in 2009 for primary science, although it remained for the other core subjects of literacy and numeracy. And, while the abolition of national testing in science may have led to a broader and more balanced science curriculum in some schools (Wellcome Trust, 2011), the diminished relative status of primary science led to a reduction in resourcing for the subject in others, with 40% of surveyed schools reporting static or decreasing budgets (SCORE, 2013), and less subject-specific professional development for teachers and Science Subject Leaders (SSLs) alike (Wellcome Trust, 2014).

A further government overhaul of assessment policy in 2014 (Department for Education, 2014) led to the abolition of assessment of pupil progress by levelling. Teachers were now required to reach a summative judgement of pupil progress based on a range of data sources, which might *include* written tests, but *not be limited* to them (Commission on Assessment Without Levels, 2015). This presented an opportunity and a risk to all primary schools: the opportunity to develop an assessment framework that produced a well-rounded summative teacher judgement of progress, and the risk that, without any central guidance on how to do this, the bespoke assessment frameworks developed by schools would not be as rigorous, reliable or manageable as the written tests used previously (Earle, 2017).

Figure 1: The TAPS pyramid.



The role of the TAPS pyramid

The TAPS pyramid is a tool to help teachers and school leaders understand how rich formative assessment data can be collected and used for summative judgement processes (Davies *et al*, 2014) and is based upon an existing model for the flow of assessment data through a school (Nuffield Foundation, 2012). The TAPS pyramid builds upon this model by specifying the types of assessment activities that would be appropriate at each level, from collecting formative data in the classroom to using it to form summative reports of pupil progress (see Figure 1). As such, it provides individual teachers and schools with a tool for evaluating their existing assessment practice and taking steps to improve it, while also exemplifying an assessment framework that fits the current English assessment policy of using teacher judgment to define pupil progress.

The TAPS pyramid has been presented at conferences, seminars and meetings of science subject specialists, and downloaded many times in the UK and abroad (Hopwood-Stephens, 2017). But how exactly has it been used, and by whom? And what impact has it had upon the assessment practice of the primary teachers using it?

Methodology

This research analyses an excerpt of data from an online survey to discover where and how the TAPS pyramid had been used in schools.

Dissemination

The TAPS pyramid user survey was hosted on a third party website and was live between December 2016 and February 2017. A link to the survey was disseminated through the website of the Primary Science Teaching Trust (PSTT) and their College Fellows network. It was also disseminated to schools applying for the Primary Science Quality Mark (PSQM), and by contacting people who had attended professional development events where the TAPS pyramid had been presented.

Design

The thirteen ongoing formative assessment activities specified in the blue layers of the TAPS pyramid (see Figure 1) were rationalised and presented as nine statements. These rationalised statements of assessment activities were reviewed by an expert panel and pilot tested before inclusion in the survey. The statements are listed in Table 1 for reference. Survey participants rated their engagement with each of the assessment activity statements by choosing from three possible responses: *I was doing this already; I do this as a result of TAPS pyramid; I don't do this yet.*

Respondents were also asked to select their most senior current job role from *Teaching Assistant; Class Teacher; Science Subject Lead; and Assistant / Deputy / Headteacher* (henceforth referred to as Leadership). They also indicated how long they had worked in primary school teaching, from the

Table 1: Rationalised assessment activities in the online survey, taken from the TAPS pyramid.

Key	Assessment activity
A	I plan opportunities for eliciting children's science knowledge and skills
B	I discuss the learning objectives and success criteria for science lessons with my class
C	I gather formative assessment data from observations, questioning and / or discussion
D	I gather evidence from a range of different science activities for assessment
E	I use formative assessment to adapt the pace and challenge of science lessons
F	I give children written or oral feedback on how to improve
G	I give the children time to reflect upon their science work
H	I judge pupil progress in science by looking at a range of formative data
I	I have a manageable system for keeping and using formative data

following groups: *0-3 years; 4-7 years; 8-13 years; 14-19 years; 20 or more years.* The groupings for years in teaching were based upon changes to national assessment policy and curriculum guidance, hence their irregularity.

Procedure and ethics

The survey took between five to ten minutes to complete. In line with British Education Research Association ethical guidelines (BERA, 2011), the purpose of the survey was made clear to participants on the first page, as well as how the data would be used. Participation was voluntary and participants could leave the survey at any time without completing it. The last page of the survey also gave the contact details for the researcher, in the event that the participants had questions or wished to withdraw their data. No requests to withdraw were received.

Analysis of responses

Once incomplete data sets were removed, the data contained 96 complete sets of responses. Descriptive statistics were generated using third party survey analysis software, with all percentages rounded to the nearest whole number.

Results

The following section describes the overall results, and the results when grouped by job role and years' experience in teaching.

Overall impact upon individual practice

Figure 2 shows the percentage of respondents stating that they now use the assessment activities listed in Table 1, as a result of their use of the TAPS pyramid.

The data show that, overall, as a result of engagement with the TAPS pyramid, activity has increased across the specified range of assessment activities. This is most obvious for assessment activity G, *I give the children time to reflect upon their science work*, with almost half of the respondents indicating that they now do this as a result of using the TAPS pyramid. Forty-two percent also report that they now *judge pupil progress in science by looking at a range of formative data* (H). This is triangulated by the finding that over one third also report that they now *gather evidence from a wide range of different science activities for assessment* (D).

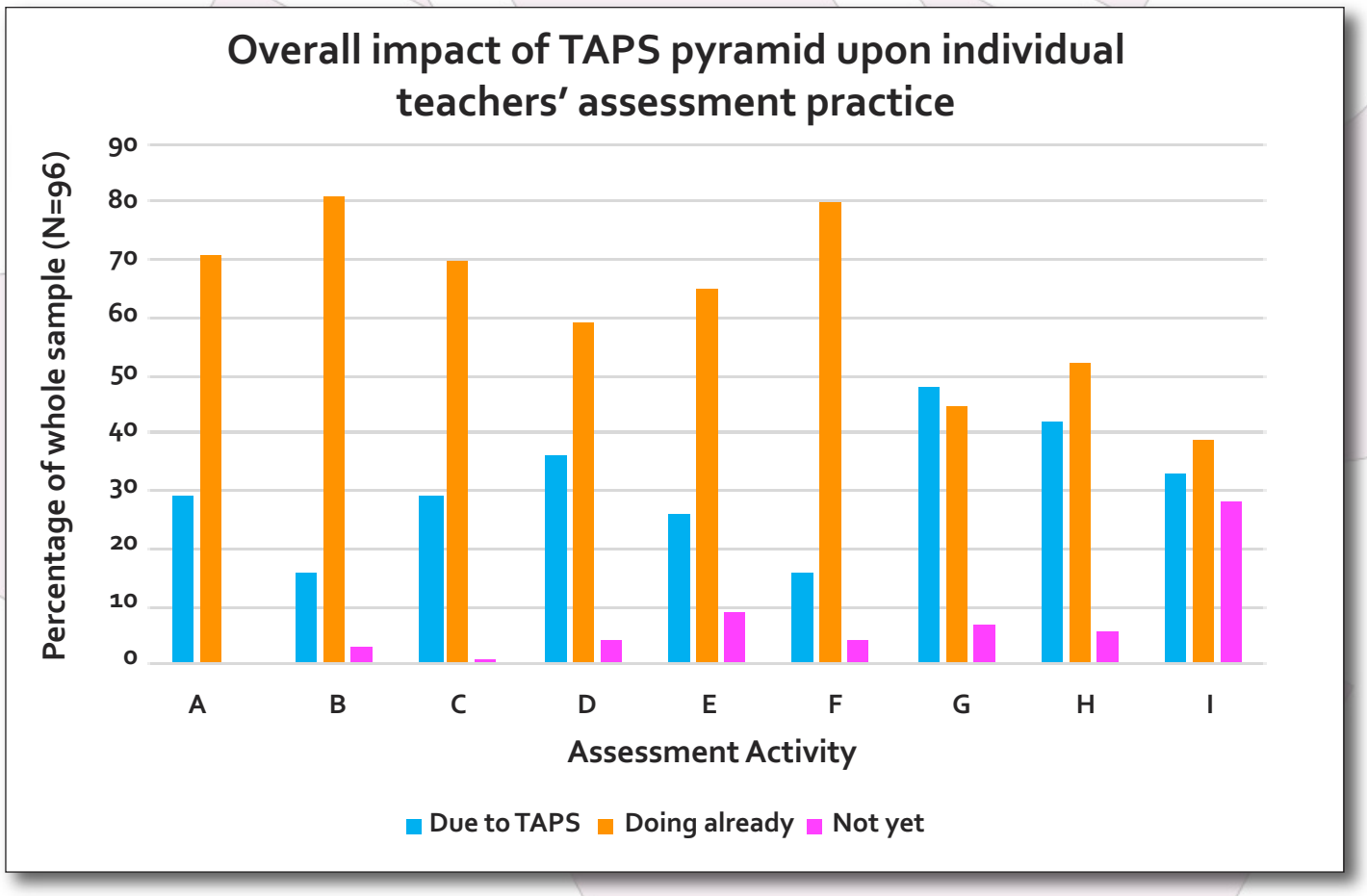


Figure 2: Overall impact of using the TAPS pyramid upon teacher assessment practice.

Taken as a whole sample, the TAPS pyramid seems to have had the least impact upon the following assessment activities: *I discuss learning objectives and success criteria for science lessons with my* class (B) and *I give children written or oral feedback on how to improve* (F). Eighty percent of all respondents indicate that they already engage in these assessment activities, possibly because they are already firmly embedded in lesson planning templates and school marking policies in many schools.

The activity that seems hardest overall for individual teachers to implement is I: *I have a manageable system for keeping and using formative data,* with 28% overall saying that they have not yet engaged with this assessment activity. Interestingly, this activity also has the lowest number of respondents indicating that they did this already.

Impact upon practice by job role

When respondents were grouped by job role, there were 12 class teachers, 73 SSLs and 11 in leadership positions.

It is clear from the graph in Figure 3 that the TAPS pyramid had the most pronounced influence upon the assessment practice of class teachers, followed by SSLs and then leadership. Sixty percent of the class teachers surveyed indicate that they now *plan opportunities for eliciting children's science knowledge and skills* (A), 60% indicate that they now *give children time to reflect upon their work* (G), and 50% report that they now *gather formative assessment data from observations, questioning and / or discussion* (C). Just under one third also report that they now *give written or oral feedback on how to improve* (F).

In comparison, SSLs are more likely to already be engaging in those assessment activities. Instead, they are more likely to report that they now *gather evidence from a range of different science activities for assessment* (D) and *use formative assessment to adapt the pace and challenge of science lessons* (E) as a result of using the TAPS pyramid.

Among leadership roles, the TAPS pyramid has had most impact upon *giving the children time to reflect upon their work* (G). Where it had no impact upon practice (activities F and I), it was due to respondents stating that they already engaged in those activities. The modest impact upon assessment activities such as *using formative assessment to adapt the pace and challenge of*

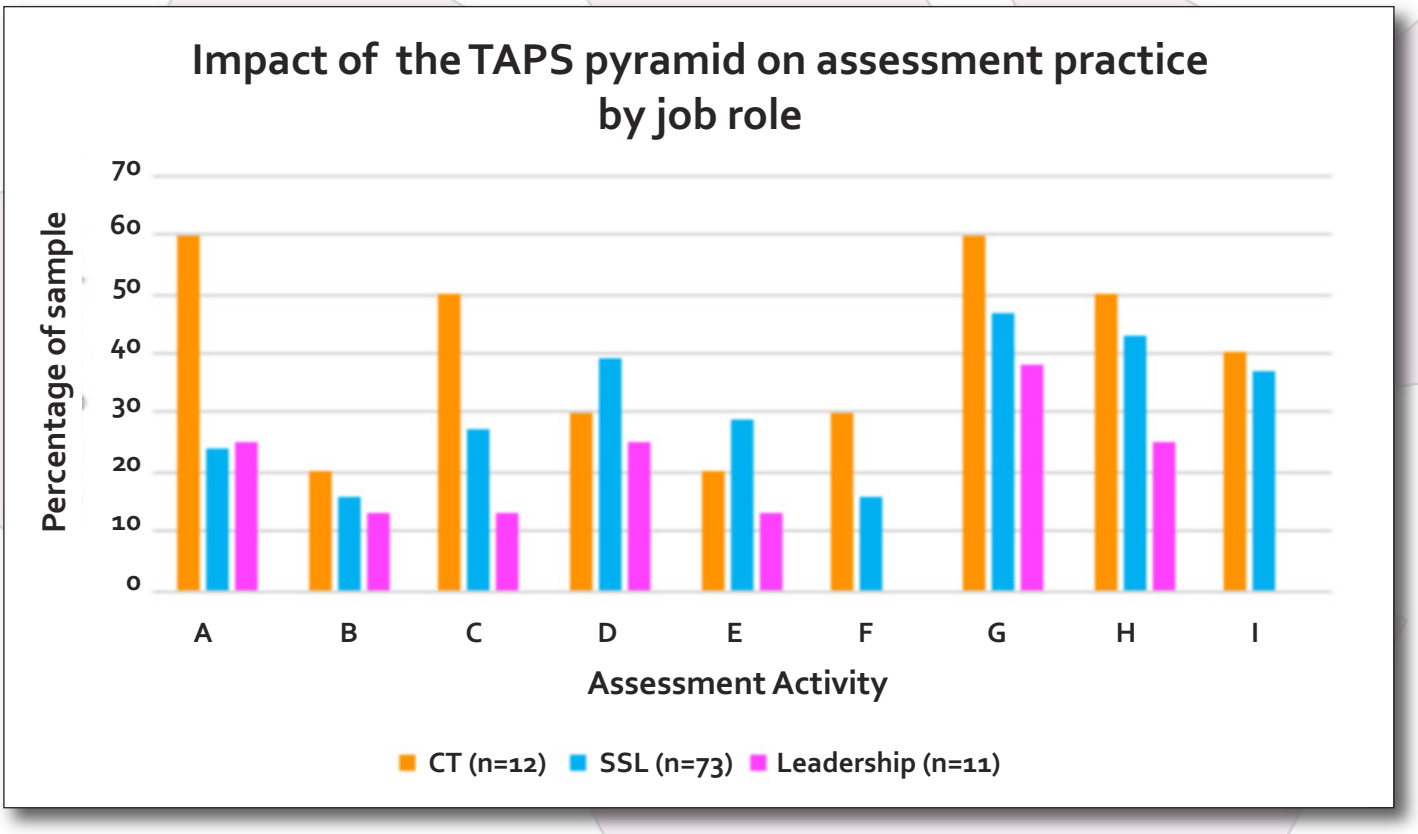


Figure 3: Graph to show the impact of the TAPS pyramid upon assessment practice, by job role.

lessons (E) may be partly due to these respondents having no class teaching responsibilities.

In summary, the impact upon practice seems to vary according to the job role and associated responsibilities of the person using it, with class teachers using it to develop their range of formative assessment strategies and SSLs using it to expand their use of the formative data that they were already generating. It has had the least impact upon the practice of those in leadership roles.

Impact upon practice by years in teaching

The line graph in Figure 4 shows the impact of each of the assessment activities across the respondents' years in primary teaching.

A prominent feature of this graph is the peak in impact among teachers who have been teaching for three years or less. For activities C: *I gather formative assessment data from observations, questioning and / or discussion*, G: *I give the children time to reflect upon their work* and H: *I judge pupil progress by looking at a range of formative data*, 80% of this group report that they now engage in these assessment activities as a result of using the TAPS pyramid.

A second, smaller, peak in impact can be seen in some of the assessment activities for teachers who have worked for between eight to thirteen years, such as *I have a manageable system for keeping and using formative data* (I) and *I give the children time to reflect upon their work* (G).

There is a further spike in the impact upon practice for respondents who have taught in primary schools for the longest (twenty years or more); the only activities that show a decline in impact upon practice for this group are *I plan opportunities for eliciting children's science knowledge and skills* (A) and *I give children written and oral feedback on how to improve* (F).

Overall, it looks as though the TAPS pyramid has influenced assessment practice most in those teachers who are newest to the profession, followed by those who have worked in it for the longest.

Discussion

The following themes identified in the results will be explored in this section: the changing impact upon practice as job role and years in teaching change, and also the diversity of assessment literacy among the primary workforce.

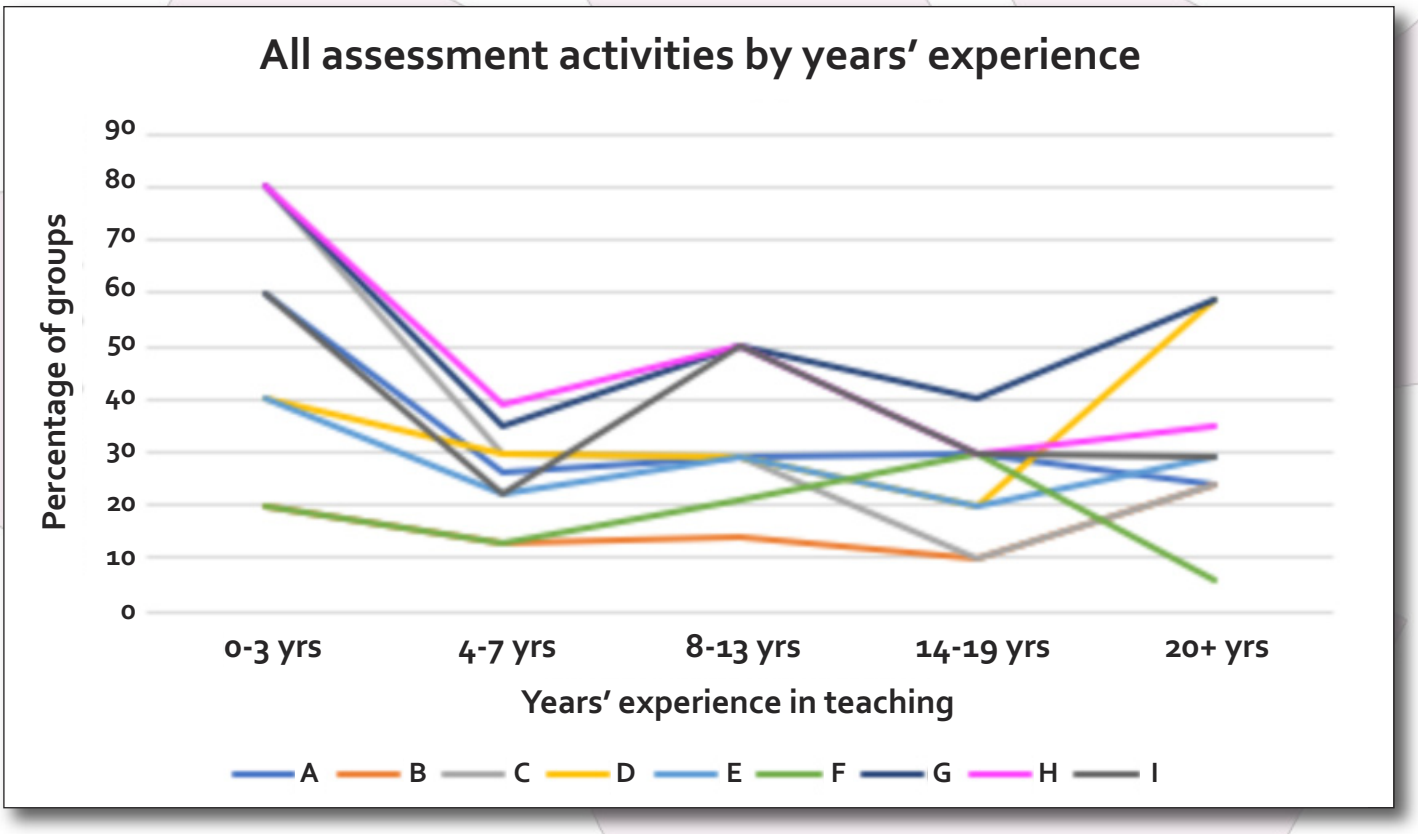


Figure 4: Line graph to show impact of TAPS pyramid across years in teaching.

Impact upon practice changes with seniority of job role

These data show a noticeably higher level of impact upon the assessment practice of class teachers, compared to SSLs or leadership roles. This might be partly explained by a lower baseline level of assessment literacy among this group but, if so, what has caused this?

As previously discussed, primary science has been demoted in status from a core subject with regional training centres and staff development budgets (Harlen, 2008) to a lower priority core subject with no national summative testing programme for all pupils. Reductions to school funding since 2010 have also severely curtailed school budgets for training and professional development (Teacher Development Trust, 2017) and the increased autonomy of schools to set their curriculum leaves primary science vulnerable to the preferences and priorities of school leaders (Ickowitz-Seidler, 2017). Class teachers working within this context may not be given access to the subject-specific training in aspects of primary science that are offered to their subject-leading peers, nor be aware of local subject support networks for primary science. A possible explanation for the study results is that, due to constrained resources and science's deprioritised status in the primary curriculum, class teachers are a neglected group within the school for receiving professional development in this subject.

This notion is lent further weight by analysing the impact of the TAPS pyramid upon the assessment practice of SSLs. This group reported that they were already using many of the assessment activities that had such an impact upon class teachers' assessment practice. By contrast, the biggest impact upon the assessment practice of SSLs was upon activities that put the rich formative data that they were already gathering towards further use, such as adapting the pace of the lesson as they taught it or forming summative judgements of pupil progress from a range of formative data.

The limited impact of the TAPS pyramid upon the assessment practice of the leadership group might be partly explained by the changes in assessment policy that they have worked under – 45% of this group had served in teaching for twenty years or more – but also the reduced likelihood that they have a regular teaching responsibility.

The changing influence upon practice as seniority of role increases can be seen as further proof that there is no 'one size fits all' in the professional development of a teacher workforce (Hargreaves, 1999). As such, it serves as a reminder to school leaders that their perceptions of the training needs and priorities in their school should not be defined by them alone, but in partnership with those teaching in the classrooms.

Impact upon assessment practice is influenced by years' experience of teaching

It has been argued above that one reason for a low level of assessment literacy among class teachers might be their limited access to subject-specific professional development and training. Another reason might be the adequacy of initial teacher training in the appropriate and beneficial use of formative assessment in primary science, however.

As mentioned earlier, variability of initial teacher training in the theoretical and technical aspects of assessment has been identified as an area for improvement among teacher training providers (Carter Review, 2015). The fact that so many recently qualified survey respondents indicated that their use of the specified assessment activities was due to using the TAPS pyramid may indicate that these respondents had not learned these skills – or realised their applicability to primary science – until they engaged with the resource. And while issues in initial teacher training in assessment practice for primary science might be partly explained in England by science's fluctuating status, it is worth noting that variability in developing teachers' assessment literacy is not limited to one country; this is a live international concern, which has also been explored in Holland (Heitink *et al*, 2015), Norway (Smith, 2011) and Thailand (Yamtim & Wogwanich, 2014).

In the mid-range of experience (eight to thirteen years' teaching experience), there is a smaller spike in impact upon assessment practice. Teachers in this group would have entered the primary workforce while levelling and best-fit statements for pupil progress were being used.

Now that assessment policy has shifted to teacher judgements and post-levels assessment frameworks, the TAPS pyramid might prove a useful resource for developing their repertoire of

formative assessment strategies and their use of that formative data; indeed, 50% of this group say that, as a result of using the TAPS pyramid, they now judge pupil progress by looking at a range of formative data.

At the other end of the spectrum, the TAPS pyramid has had a noticeable impact upon some of the assessment activities for respondents with twenty years' or more teaching experience. Fifty-nine percent of teachers in this group report using evidence from a wide range of activities for assessment as a result of using the TAPS pyramid.

It would be interesting to know whether the recent changes in assessment policy outlined above have given these teachers the opportunity to dust off previously learned skills for forming a teacher judgement of progress that fell from favour during the era of levelling and best-fit statements, or whether they feel that they have learned these assessment skills anew.

Further qualitative inquiry is planned to tease out the complex reasons for how and why the teachers in this study chose to engage with the TAPS pyramid. It nevertheless remains clear from these data that the TAPS pyramid is a resource that can be adapted to the needs of the teacher engaging with it. As such, it can be considered a useful and well targeted tool for teachers wishing to improve their individual assessment practice in primary science.

Supporting diverse assessment literacy in the teacher workforce

The results of this study indicate that the TAPS pyramid has helped teaching staff in various job roles and with differing experience to evaluate and improve their assessment practice. The following quote from a survey respondent, however, illustrates the opportunities and limitations of the TAPS pyramid as a resource for professional development: *'I would like to use the TAPS pyramid better, but changing practice takes time. I'm not dissatisfied with the TAPS pyramid – I think it's great – but with my current usage of it.'*

This respondent, a Science Subject Lead, is keenly aware of the limitations on her practice following her engagement with the TAPS pyramid, but seems unable to put her desire to improve into action.

It must be remembered at this point that the respondents to this survey were a self-selecting sample, reached through primary science communication networks. If one of these respondents, with access to subject-specific support and training in their role as SSL, has found it hard to know how to implement more of the activities on the TAPS pyramid, it can be assumed that those without subject-specific training and support would also struggle. As such, this quote simultaneously represents the usefulness of the TAPS pyramid as a roadmap for improvement, and its insufficiency in providing detailed directions.

This does not indicate a shortcoming of the TAPS pyramid as a resource, however. Instead, it illustrates the need among teachers for ongoing mutual support to achieve lasting and sustainable changes to their practice (Gassenheimer, 2013). As the respondent states, changing practice takes time, and many interventions to develop formative assessment skills in primary science have run over several months (Hondrich *et al,* 2016; Serret *et al,* 2017). Faced with shrinking training budgets and changing assessment requirements, the TAPS pyramid represents a tool for school leaders and SSLs to provide bespoke professional development in assessment practice to their non-specialist teaching staff. But this provision depends in turn on their own assessment competency and understanding of the need for support. In their review of the prerequisites for implementing formative assessment in Dutch primary schools, Heitink *et al* (2015) underlined the importance of a supportive work culture that facilitates the teachers' learning and, in Thailand, Yamtim and Wogwanich (2014) noted primary teachers' preference for collaborative working and teamwork to develop their assessment literacy. Perhaps this is the missing piece in the puzzle of changing practice: if the aim is to transform the practice of not just some but all of our teachers, we need to provide not just the physical resources, but also ongoing peer support for those who cannot access and engage with those resources independently.

Limitations to the study

The high number of respondents holding the role of SSL means that the experience of class teachers, while present in the data, is under-represented by comparison. Splitting the responses by years'

experience in teaching also produced some variability in group sizes. As such, all results should be viewed as indicative.

The quantitative analysis in this report has produced a useful snapshot of the impact of the TAPS pyramid upon assessment practice, but the nuanced explanations of why respondents implemented different activities cannot be discovered by this means. In the next phase of this research, case study data from schools using the TAPS pyramid will lead to a fuller understanding of the contextual, social and hierarchical factors that can affect the decisions of those attempting to improve their science assessment practice within a primary school environment.

The issue of variability in initial teacher training for assessment skills in primary science has been raised in this analysis. Although beyond the scope of this report, it would be a profitable avenue for further scholarly inquiry.

Conclusion

Baseline variations in primary teachers' training and experience of using formative assessment have created diversity in the ability of the primary teacher workforce to assess pupil progress in primary science. Teachers' timely access to relevant professional development in this area can be influenced by factors such as job role, changes to assessment policy and the fluctuating status of the subject. The impact of the TAPS pyramid upon teachers' assessment practice indicates that this resource is well targeted and useful but, if school leaders wish to use it to develop the assessment literacy of their staff in primary science, they will need to formatively assess the range of assessment skills present in their workforce before devising an appropriate intervention, because these data suggest a wide diversity among in-service primary teachers.

Acknowledgements

The author would like to thank Melissa Nice at the Primary Science Quality Mark and Dr. Sophie Franklin at the Primary Science Teaching Trust for their assistance with disseminating the online survey.

References

BERA (2011) *Ethical guidelines for educational research*. London: British Education Research Association

Black, P. & and Harrison, C. (2004) *Science inside the black box: Assessment for learning in the science classroom.* London: GL Assessment

Black, P. & Wiliam, D. (1998a) 'Assessment and classroom learning', *Assessment in Education*, (5), 7–71

Black, P. & Wiliam, D. (2009) 'Developing the theory of formative assessment', *Educational Assessment, Evaluation and Accountability,* **21,** (1), 5–31

Carter, S.A. (2015) *Carter review of initial teacher training (ITT).* London: Department for Education

Commission on Assessment without Levels (2015) *Final report of the commission on assessment without levels.* London: Crown Copyright

Davies, D., Collier, C., Earle, S., Howe, A. & and McMahon, K. (2014) *Approaches to science assessment in English primary schools*. Bristol: Primary Science Teaching Trust

Department for Education (2014) *National curriculum and assessment from 2014: Information for schools*. London: Crown Copyright

Earle, S. (2017) 'The challenge of balancing key principles in teacher assessment', *Journal of Emergent Science*, (12), 41–47

Gassenheimer, C. (2013) 'Best practice for spreading innovation: Let the practitioners do it', *Kappan Magazine*, **95,** (3)

Hargreaves, D.H. (1999) 'The knowledge-creating school', *British Journal of Educational Studies*, **47,** (2), 122–144

Harlen, W. (2008) 'Science as a key component of the primary curriculum: A rationale with policy implications', *Perspectives on Education 1 (Primary Science)*, 4–18

Heitink, M., Van der Kleij, F., Veldkamp, B., Schildkamp, K. & Kippers, W. (2015) 'A systematic review of prerequisites for implementing assessment for learning in classroom practice', *Educational Research Review*, (17), 50–62

Hodgson, C. & and Pyle, K. (2010) *A literature review of assessment for learning in science*. National Foundation for Educational Research

Hodgson, C., Pyle, K. & Shamsan, Y. (2009) *Assessment for learning in science: What goes on in the primary schools in England?* Conference poster. AEA Europe Conference, USA

Hondrich, A., Hertel, S., Adl-Amini, K. & and Klieme, E. (2015) 'Implementing curriculum-embedded formative assessment in primary school science classrooms', *Assessment in Education: Principles, Policy & Practice*, **23,** (3), 353–376

Hopwood-Stephens, I. (2017) *Does dissemination mode for research make a difference for reaching in-service primary teachers?* Paper presented at Teacher Education Advancement Network Conference, 11th May 2017, Aston University, Birmingham UK

Ickowitz-Seidler, L. (2017) *What is happening in primary science across the UK?* [Guest blog post]. Retrieved from: http://www.sciencecampaign.org.uk/news-media/guest-blog/what-is-happening-in-primary-science.html (Accessed 02.02.18)

Kahl, S. (2005) 'Where in the world are formative tests? Right under your nose!', *Education Week*, **25,** (4), 11

Klenowski, V. (2009) 'Assessment for learning revisited: An Asia-Pacific perspective', *Assessment in Education: Principles, Policy & Practice*, **16,** (3), 263–268

Nuffield Foundation (2012) *Developing policy, principles and practice in primary science assessment*. Report from a working group led by Professor Wynne Harlen. London: Nuffield Foundation

Ofsted (2013) *Maintaining curiosity: A survey into science education in schools.* Manchester: Crown Copyright

SCORE (2013) *Resourcing practical science in primary schools*. London: Science Community Representing Education

Serret, N., Harrison, C., Correia, C. & and Harding, J. (2017) 'Transforming assessment and teaching practices in science inquiry', *Journal of Emergent Science*, (12), 48–54

Smith, K. (2011) 'Professional development of teachers—A prerequisite for AfL to be successfully implemented in the classroom', *Studies in Educational Evaluation*, **37,** (1), 55–61

Teacher Development Trust (2017) *Twenty-one thousand teachers in school reporting no CPD budget, says TDT study*. Teacher Development Trust website at: http://tdtrust.org/press-release-21000-teachers-schools-reporting-no-cpd-budget-says-teacher-development-trust-study (Accessed 31.10.17)

Tymms, P., Bolden, D. & and Merrell, C. (2008) 'Science in English primary schools: Trends in attainment, attitudes and approaches', *Perspectives on Education 1 (Primary Science)*, 19–41

Wellcome Trust (2011) *Primary Science Survey Report*. London: Wellcome Trust

Wellcome Trust (2014) *Primary science: Is it missing out? Recommendations for reviving primary science*. London: Wellcome Trust

Yamtim, V. & Wogwanich, S. (2014) 'A study of classroom assessment literacy of primary school teachers', *Procedia: Social and Behavioral Sciences*, (116), 2998–3004

Isabel Hopwood-Stephens, Associate Lecturer, Primary Science and Doctoral Researcher at Bath Spa University.
E-mail: i.hopwood-stephens@bathspa.ac.uk