



**Earle, S. (2017) "But I've not got time for any more assessment': balancing the demands of validity and reliability', *Impact*, 1 (1), pp. 44-46.**

Official URL: <https://members.chartered.college/impact/1/1/but-i-ve-not-got-time-for-any-more-assessment-balancing-the-demands-of-validity-and-reliability>

Journal homepage: <https://members.chartered.college/impact>

## ResearchSPAcE

<http://researchspace.bathspa.ac.uk/>

This pre-published version is made available in accordance with publisher policies.

Please cite only the published version using the reference above. Your access and use of this document is based on your acceptance of the ResearchSPAcE Metadata and Data Policies, as well as applicable law:-

<https://researchspace.bathspa.ac.uk/policies.html>

Unless you accept the terms of these Policies in full, you do not have permission to download this document.

This cover sheet may not be removed from the document.

Please scroll down to view the document.

‘But I’ve not got time for any more assessment’: Balancing the demands of validity and reliability

Sarah Earle, Senior lecturer, Bath Spa University, UK

My title comes from a comment made by a teacher who attended the first development day for the Teacher Assessment in Primary Science (TAPS) project, an ongoing research programme funded by the Primary Science Teaching Trust and based at Bath Spa University, which develops support for teacher assessment by working collaboratively with teachers. This teacher equated assessment with marking, and explained that their current marking load was so high they could not consider doing more. With many competing demands and recent changes to assessment, it is time to step back from the daily rituals and question the principles behind practices that adversely affect teachers and students. This article considers how to balance the demands of validity and reliability through school self-evaluation, so that assessment can support student learning.

Research evidence on the impact of formative assessment, or assessment for learning (AfL), is strong, yet its implementation in schools has a chequered history, with some assuming it would be a quick fix and others equating it to frequent testing (Mansell et al., 2009; Wiliam, 2011). Assessment designed to support learning is not all about marking or giving a grade, it is about the whole teaching and learning process, active students and responsive teachers. This makes change in assessment practice a long-term goal. It takes time to change the way we teach, especially when it feels like we are going against the grain, in terms of marking policies or converting performance into numbers for the tracking sheet. Converting something into a number does not turn it into a valid assessment. For example, tracking software may define progress in percentages, but this is often a tracking of coverage rather than a valid assessment of an individual student’s attainment. Before slavishly completing more assessments to feed the tracking monster, it is time to pause and consider what it is that we want assessment to do.

A balancing act between validity and reliability

Any assessment is a balance between validity and reliability. While validity pulls in the direction of open and authentic assessment of the whole subject, reliability pulls in the opposite direction, with closed tasks which would have high interrater reliability.

The decisions we make in class directly relate to this balancing act: do we want a quick multiple-choice fact check (higher reliability) or an open elicitation of what the students already know (perhaps higher validity); do we want all the students to have the same test conditions (higher reliability) or provide a choice of ways to access the task (perhaps higher validity); do we focus on elements that are more easily assessed (higher reliability) or consider a fuller range of objectives (higher validity)? The answers to each question depend heavily on the context and purpose of the assessment, but it is important to be aware of the debate – there is no perfect assessment out there that is high on both reliability and validity: it is always a balance.

Moderation discussions support a shared understanding

Harlen (2007) suggests teacher assessment can be ‘reliable enough’ if it is supported by moderation. Moderation discussions do two jobs: the dialogue supports assessment

decisions about student outcomes, but, more importantly, it provides opportunities for professional learning about both the subject and assessment itself. When moderation is presented as an open discussion to develop subject and assessment literacy, it can be used to develop a shared understanding across the school. By understanding the subject well, teachers are able to assess more validly, because there is understanding of the range of the subject and, more reliably, because there is understanding of what it looks like to have ‘got there’ (Figure 1). As well as moderation discussions, such shared understanding can be supported by a mapping of curricular objectives or criteria; examples of planning or student outcomes; team planning or teaching; providing teaching assistants with key questions to ask; sharing objectives and developing success criteria with the students to support their self and peer assessment; a staff meeting on the range of ways to elicit or record student ideas. With assessment bound up in the whole teaching and learning process, the list of ways to develop a shared understanding is endless. The key point is that assessment needs to become part of each professional learning discussion, rather than something separate

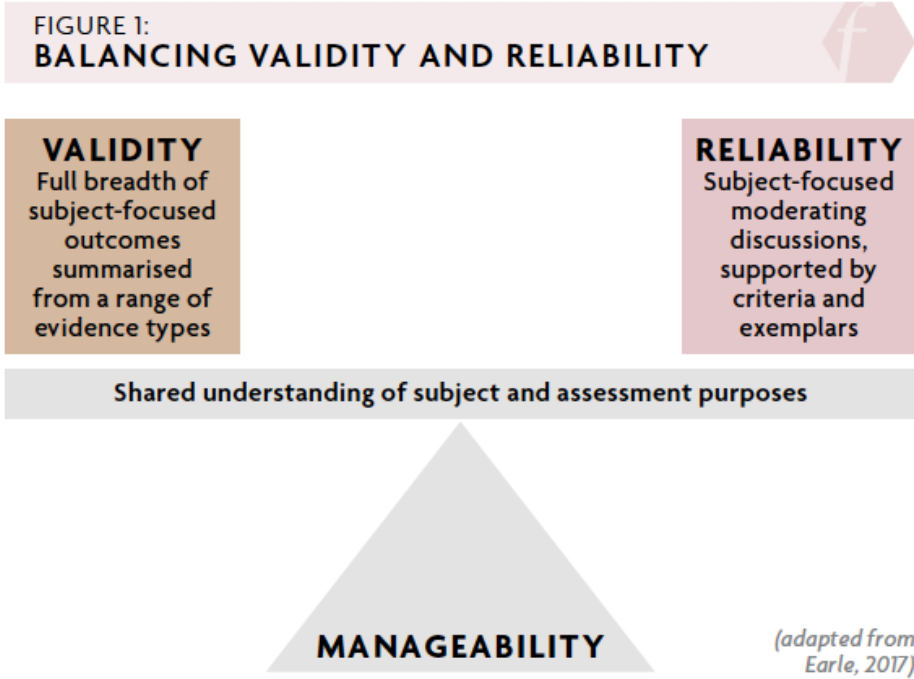
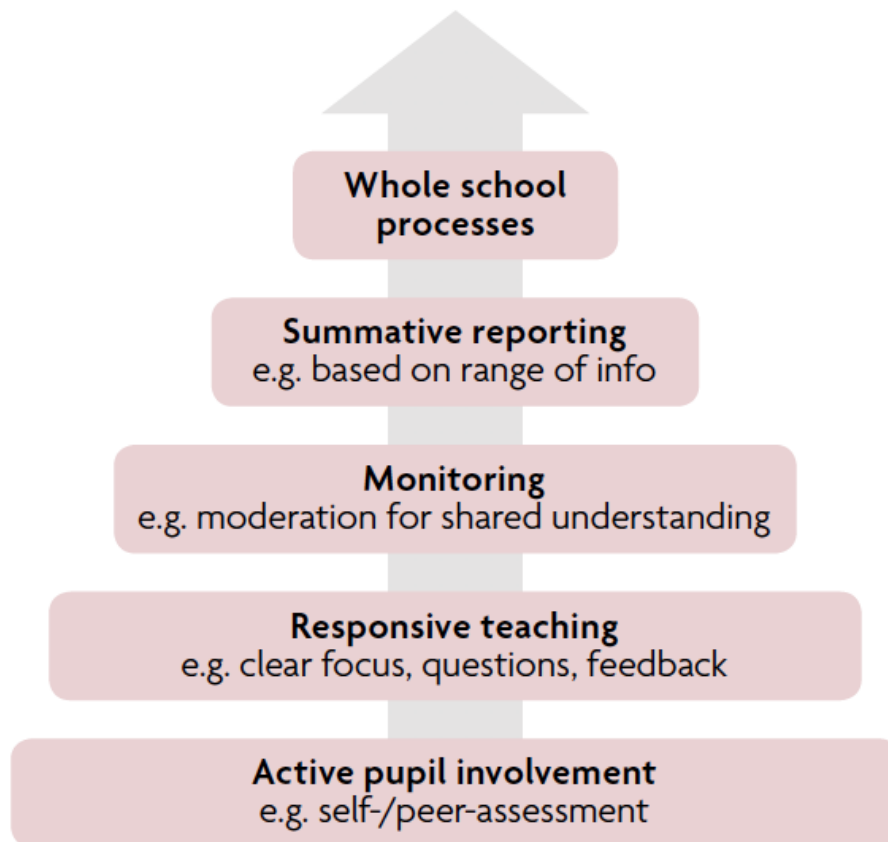


FIGURE 2:  
KEY PRINCIPLES OF TAPS



Real examples are helpful

The TAPS project has found that all of this is much more manageable if we can see examples in practice. Examples give a starting point, a way of putting this theory into practice and ideas that we can adapt for our own classrooms. TAPS has developed a bank of focused assessment plans and examples to support assessment in primary science (see website: Primary Science Teaching Trust, 2017). It has also developed a pyramid-shaped school self-evaluation tool, freely available, which contains more generic guidance about assessment. It contains a list of criteria, together with linked examples (in the context of primary science, but easily extrapolated to other phases and subjects), which schools can use to support discussion of their assessment approaches.

School self-evaluation

Self-evaluation of school assessment processes is also more manageable with examples or a supportive structure. One approach is to look at assessment practices at different 'layers' of the school, from classroom practice, to whole school processes. Figure 2 outlines the key principles of the TAPS approach, a pyramid-shaped model with indicators to support practice at each layer. The full list of criteria for school self-evaluation can be found on the website, as either a photocopyable single page or a downloadable PDF where the pyramid boxes are linked to examples. The most important layers of the TAPS pyramid are the teacher and pupil layers at the base of the pyramid, which detail how assessment can be used formatively by active students and responsive teachers. The monitoring layer is driven by middle leaders, who promote open dialogue about attainment and may point staff

towards supportive structures like progression grids. The top layers suggest that summative reporting can draw on a range of information to enhance validity, with the arrow representing this flow of information through the school.

### Classroom activities for different purposes

Not every classroom activity will, or should, lead to outcomes that would inform a summary of attainment. For example, formative elicitation activities at the beginning of a topic are more useful for adjusting planning, while summative end-of-topic posters can inform a summary but may not be useful for formative purposes if the topic will not be revisited. Teacher assessment strategies can be sorted in terms of validity and reliability (Table 1).

**TABLE 1:  
SOME TEACHER ASSESSMENT STRATEGIES**

<b>Open assessment strategies, which could be higher on validity but lower on reliability</b> <i>(good for eliciting ideas at the start of a topic)</i>	<b>Focused assessment strategies, which could balance validity and reliability demands</b> <i>(good for tracking progress across the year)</i>	<b>Closed assessment strategies, which could be higher on reliability but lower on validity</b> <i>(good for a quick check of concepts)</i>
<ul style="list-style-type: none"> <li>● Student question raising</li> <li>● KWL grid (Know, Would like to know, Learnt)</li> <li>● Mind map/thought shower</li> <li>● Student drawing</li> <li>● Concept cartoon discussion</li> <li>● Investigation</li> <li>● Student presentation</li> <li>● Observation</li> </ul>	<ul style="list-style-type: none"> <li>● Focused teacher questioning</li> <li>● Whole investigations with focused recording of one element</li> <li>● Choice of challenge tasks</li> <li>● Observation supported by a working scientifically tracking grid or expectations on planning</li> <li>● Feedback or marking focused on the objective</li> <li>● Self-/peer-assessment using success criteria</li> </ul>	<ul style="list-style-type: none"> <li>● Written tests</li> <li>● Quick-fire teacher questions</li> <li>● Multiple choice quiz</li> <li>● Close-the-gap worksheets</li> <li>● Diagrams with pre-made labels</li> <li>● Investigation recipe cards</li> <li>● A directed sorting activity (only one way to sort)</li> </ul>

*(Source: Earle, 2017)*

### Assessment choices

This article has sought to argue that a discussion about validity and reliability is relevant and worthwhile when considering how to develop assessment practices. At each 'layer' of the school, self-evaluation can be used to explore how to balance the competing demands of validity and reliability. Assessment is a key part of teaching and learning, and something that we can make decisions about on a daily basis: for example, by choosing how open or closed an activity is, or by deciding if and how to record student outcomes. If assessment can be embedded within planning and teaching, it is more likely to support student and teacher learning.

## References

- Earle S (2017) The challenge of balancing key principles in teacher assessment. *Journal of Emergent Science* 12: 41-47. Available at: [www.ase.org.uk/journals/journal-of-emergent-science](http://www.ase.org.uk/journals/journal-of-emergent-science) (accessed 23 August 2017).
- Harlen W (2007) *Assessment of Learning*. London: SAGE Publishing.
- Mansell W et al (2009) *Assessment in Schools: Fit for Purpose?* London: Teaching and Learning Research Programme. Available at: [www.aaia.org.uk/content/uploads/2010/06/Assessment-in-schools-Fit-for-Purpose-publication.pdf](http://www.aaia.org.uk/content/uploads/2010/06/Assessment-in-schools-Fit-for-Purpose-publication.pdf) (accessed 23 August 2017).
- Primary Science Teaching Trust (2017) *Teacher assessment in primary science (TAPS)*. Available at: <https://pstt.org.uk/resources/curriculum-materials/assessment> (accessed 23 August 2017).
- William D (2011) *Embedded Formative Assessment*. Bloomington, IN: Solution Tree Press.