



Earle, S. (2021) 'Principles and purposes of assessment in the classroom', *Impact*, 12, pp. 20-23.

Official URL: <https://impact.chartered.college/article/principles-and-purposes-assessment-classroom/>

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

ResearchSPAce

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

This is a pre-published version of the text.

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.

Principles and purposes of assessment in the classroom

Sarah Earle, Bath Spa University

It might not always feel like it but, as teachers, we are constantly using assessment information to make decisions about what to say, what tasks to set and what to do with task outcomes. Every interaction with students is a potential assessment opportunity, in the sense that such interactions provide us with information about how learning is going, in order to help us to adapt our teaching.

Assessment influences school and classroom culture, impacts on pupil and teacher ideas about learning and determines what is taught and how. According to Stobart (2008, p. 1), ‘Assessment does not objectively measure what is already there, but rather creates and shapes what is measured.’

The effective use of assessment provides both the means to identify whether students have succeeded and the information to help teachers to support those who have not yet ‘got there’. This article will emphasise the *use* of assessment as we consider principles and purposes. It will explore formative and summative purposes to support meaningful use of assessment, and the principles of validity and reliability to inform decision-making.

Formative and summative purposes

In its broadest sense, assessment is an integral part of teaching and includes ‘the process of gathering, interpreting and using evidence to make judgements about students’ achievements’ (Harlen, 2007, p. 11). All interactions with students potentially provide information that could support teachers to make judgements. And students should be included as judgement-makers, involving them in the active monitoring of their own learning.

The purpose of assessment is often hard to define, with information being used by a range of people for a variety of reasons. An important classification concerns formative and summative purposes. It is important to note that it is the *use*, rather than the activity, that designates the categorisation, because the majority of tasks can be utilised for formative or summative purposes. For example, a multiple-choice quiz can be used formatively to diagnose gaps in understanding or summatively to check understanding at the end of a unit.

Formative purpose in practice

Formative assessment concerns the ongoing classroom assessment practices that inform teaching. This could be something done at the beginning of a topic, to inform planning for the term, or it might be something done in lessons to check whether students need more time on a concept or whether they are ready to move on. In lessons, opportunities for formative assessment can benefit from consideration in advance – for example, deciding on key or ‘hinge-point’ questions or the focus for student recording.

Black and Wiliam (2009) identified the following aspects of formative assessment, which are considered in more detail by Wiliam (2018):

- where the learner is going – clarifying learning intentions and criteria for success
- where the learner is right now – eliciting evidence of student understanding through questioning, discussion and other learning tasks
- how to get there – providing feedback that moves the learner forward
- utilising peer and self-assessment.

These aspects could act as a guide for teachers to prompt reflection on classroom practice and help select which element could be the focus for professional development. For example, if a teacher finds that their questioning is not providing useful information about student understanding, then they may explore ways in which to increase the ‘wait time’ to provide the opportunity for more in-depth discussion and more detailed replies (Black et al., 2004). The aspects listed above are general principles rather than specifics for each lesson because formative assessment is not a list of strategies or a ‘recipe’ to follow; it requires ongoing reflection within and about the lesson. Responsive teachers should utilise their pedagogical content knowledge (PCK) to develop and refine interactions with their students to support learning.

Pupils give feedback about their learning in each verbal or written interaction. When you have identified a need, a gap or a misconception, then the key to formative assessment is to make sure that you do something with this information. For example, you might:

- ask the question in a different way to support understanding
- provide an additional explanation or demonstration
- make a note of a tricky concept to address in a later lesson
- identify those pupils who need some extra support in a particular area
- give verbal feedback to be acted upon in the lesson
- direct pupils to the agreed success criteria to support their self/peer assessment.

Summative purpose in practice

Summative assessment might be based on a ‘snapshot’ – an activity at a particular point in time, like an end-of-term test, or a summary that takes a range of information into account, like an end-of-key-stage teacher assessment (Earle, 2019). Utilising a wide range of information when drawing conclusions ‘by looking at patterns of performance over a number of assessments’ (DfE, 2019, p. 19) can help us to have more confidence in our judgements, because we are less likely to be focusing on results that are context- or task-dependent. Nevertheless, if all of the assessment tasks are drawn from the same pool, then we may still want to consider how much trust we place in our judgements. For example, written assessments for young children where many of the class cannot yet read fluently may only tell us about reading attainment, rather than knowledge of the topic.

Decisions about purposeful summative assessment should be directly related to the primary aim of reporting or summarising attainment. With this in mind, it is useful to consider the audience: who is the report or summary for? Identifying the audience will help to decide the amount of detail and the language used, since this will vary depending on whether the report is for a pupil, other staff (next teacher, head of department, etc), parents, etc. Nevertheless, assessment that has a primarily summative purpose can still be used in a formative way – for example, to identify gaps to inform the next term’s planning. Teachers may use summative assessment information over a longer period of time to support the development of their practice or the school’s curriculum.

The competing uses of assessment can place the teacher in a ‘conflicted position’ (Green and Oates, 2009, p. 233). Assessment for accountability may seemingly require a different approach to using assessment as part of the learning process. When feeling this ‘conflict’, a discussion with colleagues could be useful to clarify the purpose of the assessment. Teacher assessment literacy is an ongoing and developmental process (DeLuca et al., 2016) and such collaborative reflection can be useful for all colleagues.

The principles of validity and reliability to inform your decision-making

When making decisions about what assessment task to do or what information to gather, a consideration of the principles of validity and reliability can be useful. Both validity and reliability can be examined in great depth; to keep our discussion focused on the classroom, a brief definition for each is presented here, before discussing each in turn:

- validity: whether an assessment is fit for purpose and actually assesses what we want it to – does it merit the inferences that we base on it?
- reliability: trust in accuracy or consistency of an assessment.

Validity

Construct validity concerns how well the assessment samples the underlying skill, concept or subject (Stobart, 2009). When deciding on an assessment activity, it is important to consider what you would like to know: which knowledge, understanding or skills should be the focus? Recognising that an assessment activity can only sample a small part of the curriculum, it is worth confirming which part you are wanting to know about (not just the part that is easy to check!). This will help to decide whether the task is fit for your purpose (Green and Oates, 2009) and whether your inferences based on the results are justified. For example, a times tables test can support inferences about a child’s recall of multiplication facts, but not about the child’s attainment in mathematical problem-solving. Checking whether our inferences are justified helps to challenge our preconceived assumptions about our students. We all have preconceived assumptions, which help to make our teaching more manageable – we plan our lessons by second-guessing what students will be able to do – but assessment helps us to check whether they were in fact able to do it or not.

Two threats to validity are useful to consider when exploring the validity of assessment judgements. ‘Construct underrepresentation’ is the name given to issues of limited sampling of the subject, when the assessment is only focused on a small part of what you are interested in. For example, if only the decoding of words is assessed for reading, comprehension of the text will be underrepresented. To alleviate this threat, you need to either broaden the assessment information (broader task or utilising more tasks over time) or limit your inferences to more limited judgements about the sample assessed. The second threat to validity is ‘construct irrelevance’, whereby something gets in the way of the thing that you are trying to assess – for example, if the maths questions were too hard to read or marking is focused more on the neatness of the handwriting than the historical enquiry skills that were the focus for the assessment. Being very clear about the objective(s) being assessed helps to alleviate this threat to validity.

This discussion links to the question of whether the assessment is considering learning or merely performance – for example, has the pupil said the right word to get the mark, even if they have not understood it? One student supplying a correct answer for the class may be a

‘poor proxy for learning’ (Coe, 2013, p. 12). We can only assess the behaviours that we see, so we are reliant on performance to a certain extent. However, by drawing on a range of information and by discussing and questioning further, we can be more confident in our judgements. For example, if pupils use the right word, does that mean that they understand? You may need to question further or ask them to explain. Do they need to ‘say it’ just once? You may need to ask them to demonstrate their learning on more than one occasion, e.g. revisiting the topic later in the term.

Reliability

Reliability concerns the trust that we have in the accuracy or consistency of an assessment (Mansell et al., 2009) – for example, whether we would expect a similar result if we had asked the questions on a different day, or whether we trust the assessment enough to be able to compare between groups (if we need to). This is not just an afterthought; if we do not need to compare with other groups, in particular when we are using assessment formatively, then reliability is less of a concern. If the assessment is primarily about supporting students’ learning, then sitting the task in comparable conditions, etc., is not a priority. Reliability should be more of a concern for assessments with a primarily summative purpose.

Reliability issues can be split into internal and external. Internal reliability concerns the task itself – for example, whether the wording of questions is clear enough to mean the same to everyone, since there may be terms that are reliant on previous knowledge, which could disadvantage some. External reliability concerns issues outside of the task, such as marker consistency, which concerns whether other markers agree with your judgement. In situations where it is important to reach agreement, lists of criteria or mark schemes might be developed. These can help markers to be consistent, but they may also narrow the indicators, to a point where the assessment is more about ticking boxes than student attainment. For example, Key Stage 2 English writing assessment tick lists arguably led to a focus on grammatical devices rather than coherent, purposeful writing, with new methods of ‘comparative judgement’ now being explored (for example, by www.nomoremarking.com) as a holistic alternative to criteria lists.

Reliability can be strengthened by:

- clearly defined criteria, e.g. success criteria, mark schemes, National Curriculum or exam board objectives
- external materials in controlled conditions (if end-of-year/key stage assessments need to be compared across groups)
- standardisation, e.g. compare work to agree the standard
- moderation, which may include standardisation, but also includes broader discussions about what ‘meeting’ and progression look like.

A final point regarding validity and reliability is that they can appear to be at odds with each other: ‘an assessment cannot have both high validity and high reliability’ (Harlen, 2007, p. 23). It is not possible to have highly repeatable, standardised assessment that samples the whole of the subject. Reliability relies on narrowing the task to help markers agree, while validity depends on the opposite: as broad a sampling of the subject as possible. This can be seen as an ‘inevitable trade-off’ (William, 2003) or a balancing act (Earle, 2017). The aim is to be reliable enough for the purpose, hence the need to be clear about the purpose of the assessment. For example, for primarily formative assessment, the support of learning is more

important than standardised conditions, while a snapshot summative task for comparison across the cohort will need to address concerns of reliability.

Conclusion

At first glance, it may appear that we just need to ‘get on with it’ when it comes to assessment, with statutory and school structures guiding practice. But as discussed in this article, assessment is an integral part of your teaching and you can make decisions about its implementation and use on a daily basis. Assessment needs to provide value and useful information, which can be put to use to impact the learning of individuals and cohorts. Pausing for reflection on assessment practice can help us to make assessment opportunities more fruitful and our teaching more responsive.

This is a shortened version of a chapter in [The Early Career Framework Handbook](#), edited by the Chartered College of Teaching and published by SAGE.

References

- Black P, Harrison C, Lee C et al. (2004) Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan* 86(1): 8–21.
- Black P and Wiliam D (2009) Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability* 21(1): 5–31.
- Coe R (2013) Improving education: A triumph of hope over experience. Centre for Evaluation and Monitoring and Durham University. Available at: www.cem.org/attachments/publications/ImprovingEducation2013.pdf (accessed 26 March 2021).
- DeLuca C, LaPointe-McEwan D and Luhanga U (2016) Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. *Educational Assessment* 21(4): 248–266.
- Department for Education (DfE) (2019) Early Career Framework. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/913646/Early-Career_Framework.pdf (accessed 26 March 2021).
- Earle S (2017) ‘But I’ve not got time for any more assessment’: Balancing the demands of validity and reliability *Impact* 1: 44–46.
- Earle S (2019) *Assessment in the Primary Classroom: Principles and Practice*. London: Learning Matters, Sage.
- Green S and Oates T (2009) Considering the alternatives to national assessment arrangements in England: Possibilities and opportunities. *Educational Research* 51(2): 229–245.
- Harlen W (2007) *Assessment of Learning*. London: Sage.
- Mansell W, James M and the Assessment Reform Group (2009) *Assessment in Schools: Fit for Purpose?* London: Teaching and Learning Research Programme.
- Stobart G (2008) *Testing Times: The Uses and Abuses of Assessment*. London: Routledge.
- Stobart G (2009) Determining validity in national curriculum assessments. *Educational Research* 51(2): 161–179.
- Wiliam D (2003) National curriculum assessment: How to make it better. *Research Papers in Education* 18(2): 129–136.
- Wiliam D (2018) *Embedded Formative Assessment*, 2nd ed. Bloomington: Solution Tree Press.