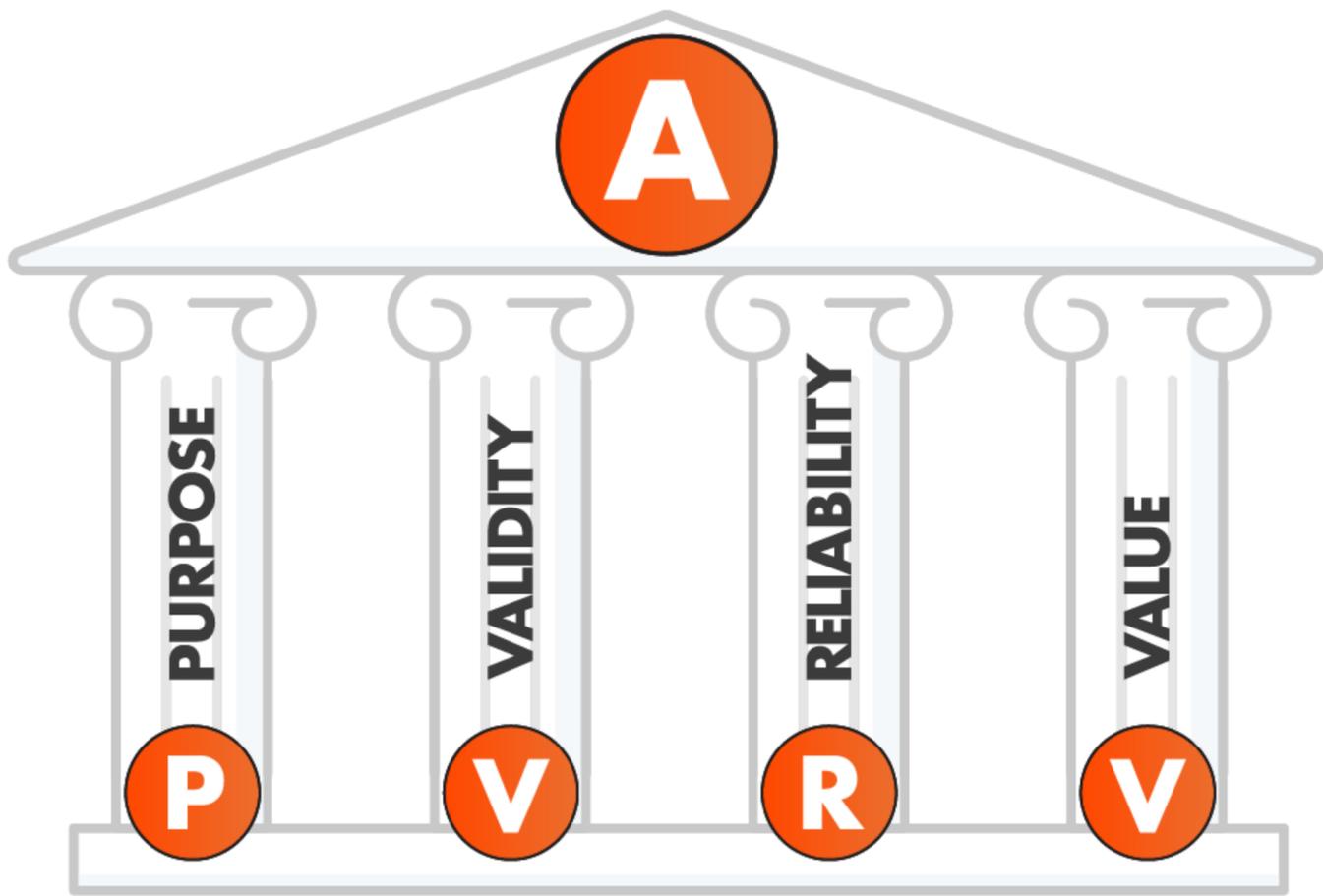


THE FOUR PILLARS OF ASSESSMENT

**A RESOURCE
GUIDE**

**BY EVIDENCE
BASED EDUCATION**





INTRODUCTION

If we were building a new school, we wouldn't start without laying firm foundations. In the same way, we can't develop great assessment practice without a strong base knowledge of the key theory around assessment.

We have distilled this theory down into the four pillars of great assessment: **purpose**, **validity**, **reliability** and **value**.

The Four Pillars of Assessment resource guide will provide you with a strong understanding of what underpins each pillar and how it supports great assessment.

Practical support to apply the principles of great assessment are available through our training courses.



1. Purpose

THE IMPORTANCE OF PURPOSE

We live in an information age, at a time when the quantity of information in our lives often outweighs the quality of it.

Assessments used to generate information on which decisions are made about student learning need to provide high-quality information fit for the purposes intended. There are dozens of reasons why you might assess pupils, and the ideal type of assessment is different depending on the purpose. However, it is not uncommon to lose sight of the function of assessment.

Assessment can also become all things to all people, where information collected for one purpose is also used as a measure of something entirely different. For example:

An end-of-year assessment designed to measure attainment in mathematics, but which tests only certain components of mathematics (multiplication, division and place values, for instance), cannot be used to draw conclusions about maths attainment in general (the intended purpose), only the components it includes. Matters are complicated when questions in the assessment contain overly complex wording. The assessment now requires sufficient reading skill to access the maths at the heart of the assessment, thereby disadvantaging weaker readers.

Until you are clear about exactly what your different purposes are, you won't be able to use the right assessments.



THREE STEPS FOR ROBUST ASSESSMENT

Purpose is the most important of the four pillars of great assessment. We must know what we want to measure and why, in order to select the right tool to achieve our purpose.

These three steps are always at the heart of any robust and purposeful assessment:

- 1. The construct:** What is the specific knowledge, skill or understanding (drawn from the curriculum) that we intend to assess?
- 2. The end use:** What do we want to do – the interpretation, the decision or action – with the information generated by the assessment process?
- 3. The best tool:** What and when is the most appropriate, effective and efficient way to assess in this instance?

Without clear answers to these questions, it is impossible to use assessment effectively. If we don't have a specific goal for an assessment, there is no way of knowing if it is any good at providing the information we need from it. And without good information, guiding students along their learning journey is difficult.

ARE YOUR ASSESSMENTS FIT FOR PURPOSE?

What sorts of assessments do you use in schools? Is everyone clear about what their intended purpose is and how the information from them will be used? Are they fit for their intended purpose, or have they been warped over time – bent out of shape to fit a need in school? Or perhaps they're done because ... well... "we've always done them"!

Make sure the most is made of your time on assessment with appropriate, dependable measures to make appropriate, dependable claims and judgements.



2. Validity

THERE IS NO SUCH THING AS A VALID ASSESSMENT!

Validity is perhaps the most commonly-used word in discussions about the quality of any assessment. While it's used a lot, it is often misunderstood and can be very misleading.

Validity is a word which, in assessment, refers to two things:

- The ability of the assessment to test what it intends to measure.
- The ability of the assessment to provide information which is both valuable and appropriate for the intended purpose

A common **misconception** about validity is that it is a property of an assessment, but in reality, there is no such thing as 'a valid assessment'.

However, there is such a thing as 'an assessment which is valid for a specific purpose': validity is all about the inferences you make based on the information generated.



In many cases, there are two reasons that assessments end up not quite hitting their target: construct under-representation and construct-irrelevant variance.

Construct under-representation: is where the assessment fails to capture important aspects of the construct, e.g., it is not fully representative of the target of the assessment.

Construct-irrelevant variance: the assessment outcomes are influenced by things other than just the construct, e.g., inaccessible language.

When we talk of validity and great assessments, we are referring to the assessment's ability to support the claims we want to make based on the information generated.

IMPROVING VALIDITY

One of the key validity checks we can do when assessing the quality of an assessment is to consider: is there either construct under-representation or construct-irrelevant variance in this assessment? Defining the construct – saying what is and isn't included in it – is a vital part of a robust assessment process. It is one way in which we can avoid construct under-representation and construct-irrelevant variance.

Ensuring that an appropriate and meaningful range of marks is used to represent performance at particular levels of achievement is another aspect of improving the validity of an assessment. If there are 50 marks available on an assessment task, but no student is awarded more than 35 marks or less than 20, is the assessment really out of 50?

Assessment validity is all about the inferences you make based on the information generated. Therefore, it is important to ask, does the assessment allow you to make inferences which are valid?



3. Reliability

WHAT IS A RELIABLE ASSESSMENT?

Reliability in the assessment of student learning is about accuracy and consistency over time and context. As we saw with validity, a determination of how reliable an assessment needs to be is informed by its intended end uses.

There are lots of factors which contribute to the reliability of an assessment, but two of the most critical for teachers to acknowledge are:

- the precision of the questions and tasks used in prompting students' responses
- the accuracy and consistency of the interpretations derived from assessment responses

"NO ASSESSMENT IS 100% RELIABLE BUT WE SHOULD UNDERSTAND RELIABILITY IN ORDER TO IMPROVE IT"

An assessment is a means by which we can create a set of circumstances in which a student can represent their knowledge, skill and understanding in an observable form. Because it is a proxy for something unseen, and because interpretation is often part of making sense of the information derived from an assessment, error is always present in some form or other. Some sources of error include:

- the assessor's unfamiliarity with the topic being assessed
- the assessor's unfamiliarity with robust assessment practices
- bias (teachers are human, after all!)
- the subjectivity of the material to be assessed
- the conditions in which students take the assessment



IMPROVING ASSESSMENT RELIABILITY

There are lots of ways in which classroom assessment practices can be improved in order to increase reliability, and one of the most immediate is to improve rater reliability.

Inter-rater reliability: getting people to agree with one another on simple matters can be hard enough, so when it comes to complex judgements (such as whether the grades two teachers award independently for the same writing task are consistent with each other), reliability challenges arise.

Intra-rater reliability: most people acknowledge that it is difficult to achieve high levels of inter-rater reliability, but an often overlooked challenge also comes from the accuracy and consistency of one's own judgements. Imagine your responses to a set of different assessment tasks of the same quality, but at different times during the day, week, month and year. Particularly in areas of subjectivity – where judgement is needed – you can imagine how your decisions, comments and grading of assignments may vary dependent on time of day, hunger, how many other tasks you're juggling in your mind, caffeine ingestion...

Improving rater reliability begins by acknowledging that assessments always have a degree of unreliability inherent in them. Improving reliability will improve the quality of the information derived from the assessment process, thus increasing its potential value to teachers and students. Below are three ways to improve reliability of assessment in school:

- Use exemplar student work to clarify what success looks like in specific assignments: be explicit about these criteria
- Blind-mark assignments: this reduces bias and increases rater reliability
- Blind-moderate samples of students' work: this increases rater reliability and also offers a good professional development opportunity to share standards

Well-designed multiple-choice quizzes can be a reliable form of assessment and can offer diagnostic information to support teaching and learning activity.



4. Value

IS IT WORTH THE EFFORT?

Every minute that a teacher, student, leader, parent or governor spends engaging with assessment is a minute that wasn't spent doing something else. As such, assessment carries with it a high opportunity cost; the value derived from the former should be at least commensurate with the latter (if it's not, we should probably do something else).

Calculate the time spent on a single assessment process (creating, administering, pupils doing, marking and feedback to pupils, data entry). How much time was taken? How much value did the assessment add to the learning process?

Our first pillar, purpose, is a key aspect of assessment value. A clear purpose increases the value of assessment information by ensuring that, as Ronseal would say, **"it does exactly what it says on the tin"**.

SURPRISE, SURPRISE

An assessment that can't surprise you isn't an assessment, so value also lies in the power to raise an eyebrow.

Building a bridge between teaching and learning is part of the role played by assessment; we can't physically peer into a brain to find out how much geographical knowledge is in there, so we need proxies. We need bridges. Yet, if the bridges we use are not built robustly, and in a way that allows new and unusual views to be encountered as we cross, they fail to function as they should and their value to teachers and learners diminishes drastically.

Do you and your teaching staff know enough about assessment to describe accurately the quality of assessments used in your school or college?



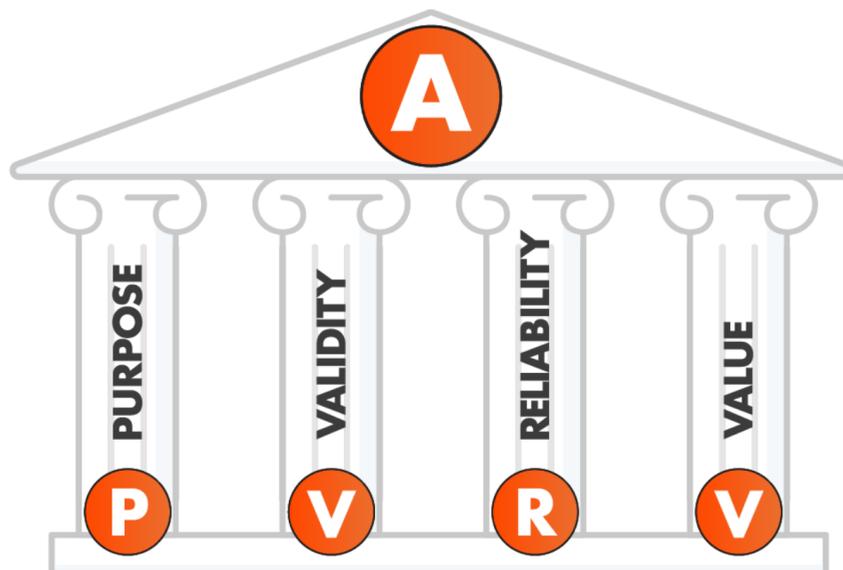
WASHBACK

Assessments can have positive and negative effects, something known as the washback effect. The intended effects of assessment, such as pupils studying more, or high-quality feedback for learning, are known as positive washback. The unintended negative effects from assessment – such as unmanageable workload, teaching to the test, decreased time for other activities – are the negative washback effects.

In many ways, effective assessment is learning how to maximise positive washback and minimise negative washback. The main way in which we approach this is to create strong and explicit links between curriculum, pedagogy and assessment.

"IT IS INFORMATION DERIVED FROM WELL-DESIGNED, PURPOSEFUL, PLANNED ASSESSMENTS WHICH BRIDGES THE GAP BETWEEN TEACHING AND LEARNING."

Better information can inform **better decisions**, and better decisions can lead to **better learning**. And if that's not the most valuable outcome, what is?





Evidence Based
Education

ASSESS BETTER

The concepts touched upon in this guide feature in the **Assessment Lead Programme** – a course of professional development to enhance the knowledge, skill and confidence to use assessment in school to support learning.

You will benefit from practical tools, guides and support so that you can effectively translate assessment theory into practice.

According to Phil Stock, Deputy Headteacher at Greenshaw High School, "the content is of the **highest class**... It draws down **all that we know about effective CPD**."

The Assessment Lead Programme delivers whole-school impact – from the improvement of your school's assessment policy and framework to the creation of powerful formative assessment questions and quizzes for use in the classroom.

To find out more, simply click on the words **ASSESS BETTER!**

