

# DESIGNING LEARNING ASSESSMENTS

---

## SUMMARY NOTE 1:

Core Concepts in Assessment

### ➔ KEY QUESTIONS:

- What can learning assessments be used for?
- How do we measure learning?
- What are reliability, validity and fairness and why are they important?



# Why Measure Learning?

## Defining specific purposes for assessment

Measuring learning is important for a variety of reasons, to a variety of stakeholders.

Generally, the purposes of learning assessments can be defined as either **summative** or **formative**.

- **Formative assessments** provide an ongoing picture of learning as it happens. This is important for teachers and other groups to help them adapt or tailor their pedagogical approaches to support a learner.
- **Summative assessments** provide an evaluative picture of whether learning has occurred over a period of time. Summative assessments are important for reflecting on the performance of an individual, system or intervention.

The purpose of an assessment will inform how it is designed, and therefore should be considered from the beginning of the design process. For Schools2030 we can think of the purpose of assessment in terms of **Results**, **Decisions** and **Impact**.

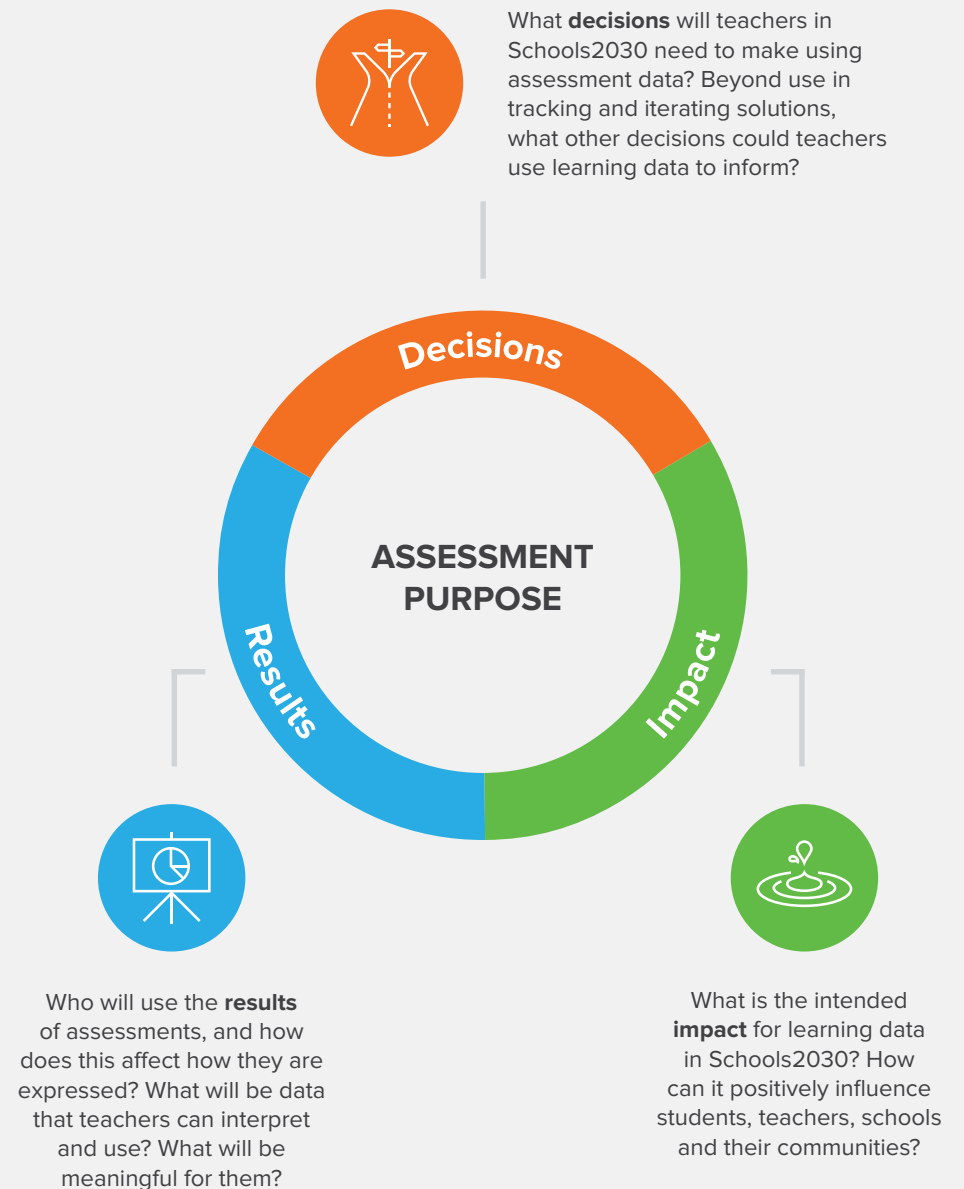


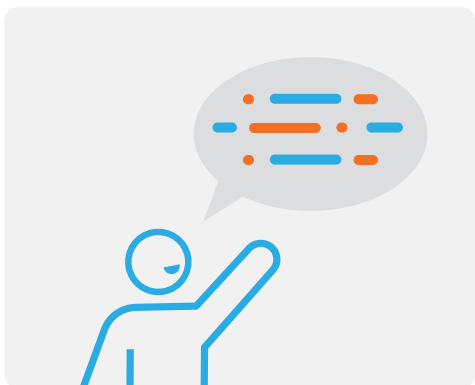
Figure 4 Defining Purpose for Assessment

# How do we measure learning?

## A case study in measuring vocabulary

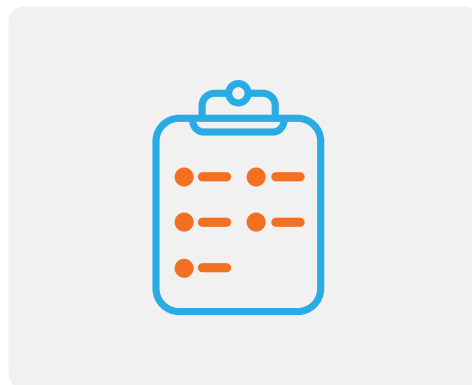
Measuring what someone knows and can do, is not a simple task.

Even simple skills contain many component parts. Trying to measure them all is often not feasible. Instead we look to measure a subset from which we can infer the total ability. **This fundamental principle of assessment is outlined here:**

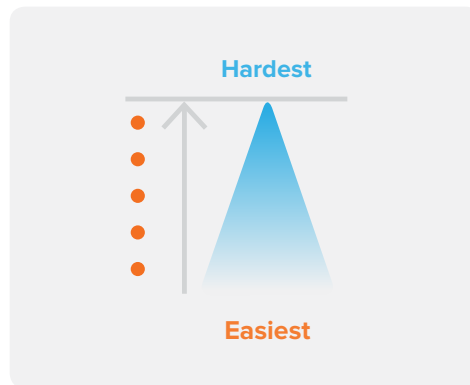


If we want to understand a child's vocabulary we face a challenge. A child may know **thousands** of words.

We can't test them on all of them to determine their vocabulary.



Instead we want to draw a shortlist from which we can **infer** a child's vocabulary



This is based on the idea that some words are easier to learn than others. Our shortlist of words will cover words that almost every child will know, as well as words that very few children will know



**We can then see how many of these key words children know.** If we have selected the right range of difficulties for our learners, we should be able to understand the total size of their vocabularies.

**This principle is applied to a range of academic and non-academic constructs to measure learning**

# Core Concepts in Learning Measurement

## Validity, Reliability and Fairness

	What is it?	Why is it important?	How to protect it?
<b>Validity</b>	The degree to which the evidence from an assessment supports the interpretations and uses of it. <b>There are two key elements:</b> (1) factors other than the measured construct affecting performance in tests; (2) failing to properly represent the construct.	Validity is the most significant quality of an assessment. It answers the question of whether or not the assessment is fit for its intended purpose.	Validity needs to be considered through all decisions at every stage of the process. The validity of a learning assessment is only as strong as the weakest stage of the process.
<b>Reliability</b>	The degree of uncertainty around observed results from assessments. That is, if a student attempts a test several times, even if no learning takes place between, will the student achieve the same score?	If reliability is poor then the scores obtained may be different from the true level of attainment. Therefore, the results cannot be relied on.	Consider all the factors that could affect how a student performs on a given day. Also consider how administration and scoring may change from time-to-time and between administrators.
<b>Fairness</b>	All students have the same opportunity to demonstrate their attainment, regardless of gender, socio-economic status, home language or any other factor.	If scores are below true levels of attainment for some groups of students then the decisions made about these students will likely not achieve their desired outcome and may be harmful.	Review how items perform differently for different students. Consider whether some students may have more difficulty responding to items for reasons other than their level of attainment.