Literacy and assessment

Assessment allows us to understand more about the progress of each individual student and about the effectiveness of different teaching approaches and interventions. Without regular, good quality assessment, it is difficult to know which teaching approaches have been more and less successful, and which students may need additional support in which areas. Importantly, regular assessment allows you to know how each student is progressing. You might not expect every student to progress at the same rate, but you do want to know that every student is improving over time.

Assessment can also reveal important information about why students are struggling with a task. For example, reading comprehension involves multiple different skills, including recognising words by sight, decoding unknown words, understanding vocabulary, punctuation and grammar, and inferential reasoning to construct the meaning of a passage of text. Different students may therefore struggle at the same reading comprehension task for multiple different reasons. For example, one student may have good language comprehension skills, but poor decoding skills, while another student might show the opposite profile. The support given to these two students will be most effective if it is targeted to their own strengths and weaknesses.

Assessment therefore has multiple purposes:

- To identify which students are doing well and which may be struggling
- To identify which students in a class have understood particular concepts
- To understand how individual students and the class as a whole are doing compared to students in other classes or schools
- To identify the strengths and weaknesses of individual students
- To measure progress over time
- To understand which approaches have been effective with which students

Factors to consider in assessment

Because assessment serves different purposes, a range of approaches to assessment is needed. The starting point is **understanding what you want to measure and why**. The next step is to decide which assessment to use and whether to use an informal or formal assessment. A wide range of factors will influence your decision-making, such as test availability, cost (in terms of time, resources and equipment), and ease of administration in your particular setting. When choosing an assessment, it is also crucial to check whether the assessment is a reliable and valid measure of the construct that you want to measure. In the case of formal assessments, the manual should provide you with the information you need to be able make objective decisions about whether an assessment is a valid and reliable measure of that construct.

Useful assessments for assessing particular skills

There are assessments available which aim to measure every different element of literacy, at the word and the text level, and there are also measures of component skills. A useful overview of educational assessment is available from Te Kete Ipurangi. Some measures of specific areas of literacy, such as

letter knowledge, reading and spelling irregular words, and reading and spelling nonwords, are freely available on the motif website.

Letter knowledge: Letter knowledge is a key skill in early literacy, and we would always recommend assessing letter knowledge when planning an intervention. In the case of assessing letter knowledge, it is normally sufficient to ask students to give the names and sounds when shown the 26 letters and some additional digraphs (such as 'ch', 'sh' or 'th'). The role of letter names and letter sounds in learning to read and write is discussed in more detail in The pedagogy of reading.

Word reading: Word reading tests can be constructed straightforwardly from lists of vocabulary you believe students might have encountered. There are also several standardised measures available. Some measures of word reading provide measures of skill on different types of word: **regular words**, **irregular words** and **nonwords**. This can be useful because it provides evidence about the types of approaches a student is using to read words. Thinking back to the dual route models of reading mentioned in Theories of early literacy development, words can be read either by decoding or by recognising them by sight. Regular words such as 'cat' or 'carpet' can be read in either way – by decoding or by recognition. Nonwords and irregular words, on the other hand, put more pressure on one pathway than another. In order to read an irregular word such as 'choir' or 'yacht' correctly, a student cannot rely on letter-sound correspondences alone. They might recognise it by sight, or guess based on partial letter cues, known oral vocabulary, and context if available. This therefore measures their reading vocabulary.

Nonsense word reading: Some measures of literacy development use 'nonsense' words or 'alien' words rather than real words. This can seem counter-intuitive because it asks students to decode words that do not exist. It is worth highlighting that nonsense words should be used mainly as an assessment approach rather than for teaching. They are a particularly pure measure of one aspect of reading (decoding). With a nonsense word such as 'dalcrit' or 'glump', the word is not in the student's oral vocabulary and so sight word recognition or guessing is not effective. Decoding is the only way to read the word successfully.

It could be argued that it does not matter how the student worked out the word provided they work it out correctly. In a typical reading situation, this is correct. The main aim is to work out the word and move on. However, in an assessment situation, we want to know more about the particular skills a student possesses because that provides useful information for the future. A student's ability to work out a particular nonsense word gives an indication of how well they can sound out any new word they have not encountered before. In other words, nonsense word reading gives an indication of how well the student has cracked the alphabetic code of literacy. We know that this ability to sound out unknown words is a powerful predictor of how well a student's reading vocabulary will grow over time. A student who has good nonword reading skills has good phonic decoding skills and is in a good position to learn new words through self-teaching (sounding out while reading independently). A student who has good irregular word reading skills has a good reading vocabulary, often indicating they have read a good range of books. Both skills are important for good literacy development.
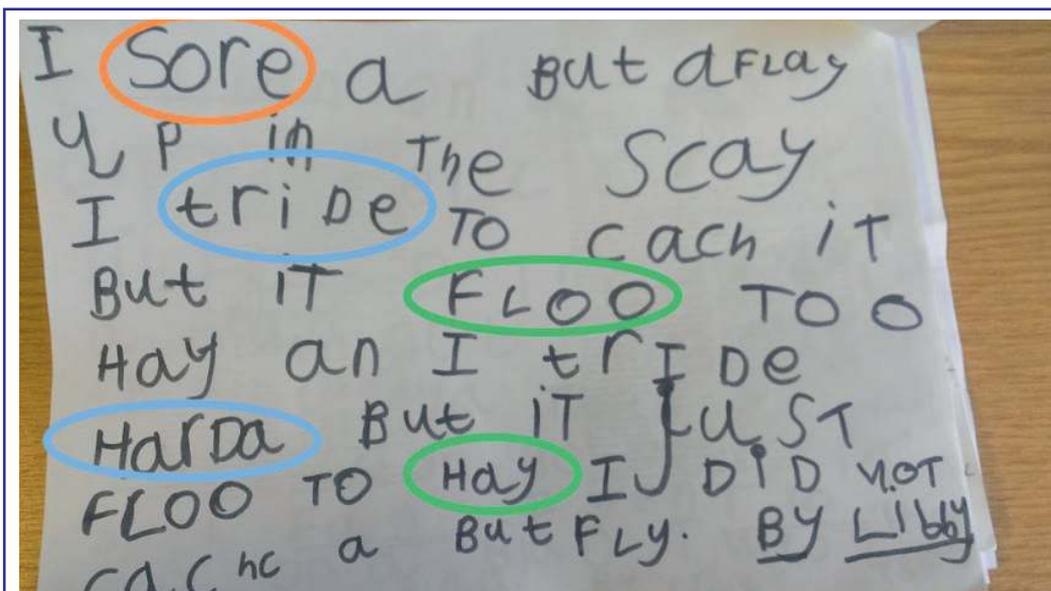
Vocabulary: Word knowledge can be tested in three ways - whether the student understands the word (for example, given a number of options, can they point to the right picture that represents the word's meaning?), whether the student can use the word appropriately (can the student name a picture that represents the word, or put the word in a meaningful sentence?), and whether the student can provide a definition of the word (a good definition might be giving an alternative word that shares the same meaning – a synonym, or a definition from which you can deduce the target word). Before beginning your intervention, compile a list of words that you are going to teach the student. These may be key words, like the examples above, or words that are used in a topic the class has been working on. Test their

knowledge of these words at the beginning of your intervention, and again at the end. Remember that knowing in terms of understanding a word does not mean the same as knowing how to use or explain a word, so it is useful to test both forms of knowledge.

Using errors to understand processes

Most assessments focus on overall accuracy on the test. However, it can also be useful to think about which individual questions the students made errors on, what type of errors they made, and why they might have made these errors. This can give hints about what underlying processes are being used to read or write. Some assessments provide guidance on interpretation of errors, sometimes referred to as diagnostic information. Even if the assessment you choose does not provide explicit guidance on interpretation of errors, it can still be useful to look at the nature of errors. In fact, it can be useful to look at the nature of errors outside of the context of an assessment as well. Errors can give you clues about areas of weakness, which will be particularly informative for planning next steps.
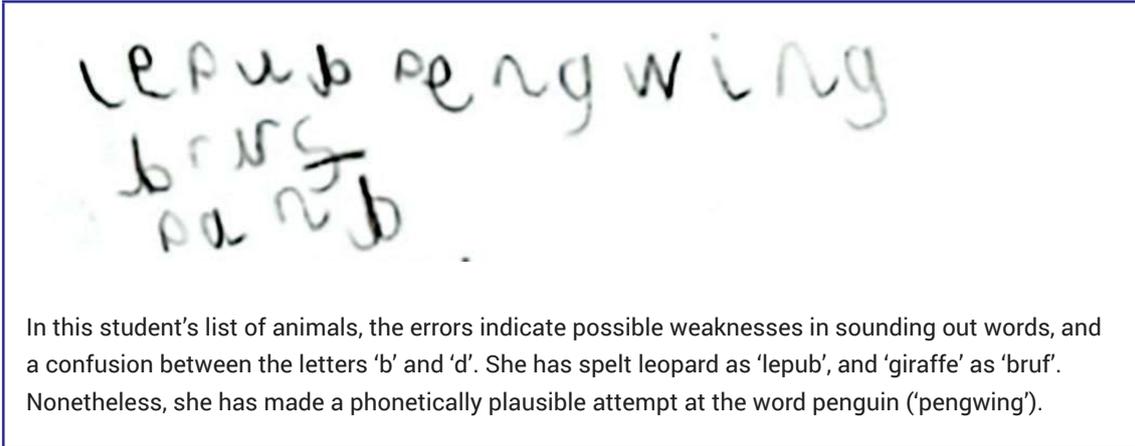
Analysing spelling errors: Looking at the types of spelling errors that a student makes might tell you something about the sources of information a student is using to produce their spellings. Are their errors consistent? Do the errors suggest that they are trying to represent the sound structure of the spoken words or not? Do they know how to spell common letter patterns such as suffixes? Figure 1 gives some examples of writing from young students to demonstrate how errors can be analysed.

This beginner speller has produced a lot of spelling errors. Looking at the types of errors they make suggests that they have good understanding of how to represent the sounds of words.

- *sore (saw) is a homophone error – the wrong word has been selected.

- *floo (flew) and *hay (high) show that the student hasn't learned the correct word specific letter combinations (orthography).

- *tride (tried) and *harda (harder) show a lack of knowledge about how to spell inflectional suffixes (word structure).

Figure 1

the education hub

In this student's list of animals, the errors indicate possible weaknesses in sounding out words, and a confusion between the letters 'b' and 'd'. She has spelt leopard as 'lepub', and 'giraffe' as 'bruf'. Nonetheless, she has made a phonetically plausible attempt at the word penguin ('pengwing').

Behavioural observations: While students are completing an assessment, it is also important to note down any behavioural observations that might be important for interpreting the results. For example, did the student become distracted or demotivated? Did they finish the assessment quickly and seem to find it easy, or were they rushed? These types of observations can be particularly informative when deciding next steps, but can also tell you something about whether the results can be relied upon. However, these observations only influence the interpretation of results - they should not actually alter the results.

Further considerations in assessing literacy

There are a number of other factors to consider when deciding what and how to assess students' progress and performance in literacy.

How speci ic is the test?

When assessing skills, some measures are very specific, while others are wide ranging. For example, a letter knowledge test assesses a very specific skill, while a standard listening comprehension test will involve a wide range of skills. Generally speaking, specific measures that are closely related to the skill you wish to improve are probably the best ways to monitor growth in that skill. For example, if you are teaching vocabulary, it is better to test improvement using a list of the words you have covered rather than a standardised vocabulary test. On the other hand, broader tests are useful for examining generalisation of the skills you have taught. For example, the ultimate aim of teaching letter knowledge and phoneme awareness is to improve reading and spelling. Therefore, it is useful to assess the impact on reading and spelling even if you have not been focusing on those skills, because improving these skills is one of the aims.

Using criterion-referenced tests versus norm-referenced tests

Many measures of progress in teaching are **criterion-referenced**. For example, within the Literacy Learning Progressions, there is a list of reading-related skills that a student is expected to show at

different stages of education. It is not the focus of these assessments to determine whether a student is average or below average for their age, but to know whether they have specific skills or not. This is **criterion-referenced** measurement. Criterion-referenced tests give lists of key skills, or, more simply, lists of words or letters that students may know at a particular stage of development.

In contrast, **norm-referenced** tests compare a student's performance to the average for their age or their year group. Tests provide information about whether a student is average, above average or below average for their age in terms of their reading accuracy and their reading comprehension. However, the test does not give information about whether the student has particular skills or pieces of knowledge.

Criterion-referenced tests are particularly important in planning teaching as they help teachers identify which knowledge and skills individual students already have, and therefore what knowledge and skills need to be specifically taught. For example, since we would expect that knowing every letter is a key skill in literacy whatever your age, it is useful to know exactly which letters a student does or does not know. In this case, a criterion-referenced test is more useful.

Norm-referenced tests are particularly important in assessing students' progress over time. For example, if an intervention carries on for six months, it is likely that over this time most students will show progress on any reading measure. However, a criterion referenced test will not necessarily show whether this progress is more or less than we would expect over the six-month period. Norm-referenced tests help us see whether a student is catching up with or falling behind in comparison to his or her peers over the period of the intervention.

Types of scores available in standardised tests

Assessment manuals sometimes provide several different ways to score performance. Assessment manuals should describe how those measures were calculated and how the results should be interpreted.

Raw scores are the sum of correct responses or points awarded on an assessment. Looking at the particular item or items with which students struggled can be useful. Raw scores are quite difficult to interpret, particularly if the questions varied in difficulty or if different students completed different questions. Because performance is influenced by both age and experience, raw scores should not be used if you want to compare or combine the results from different students. Norm-referenced assessments provide scores which have been statistically derived from a large sample of students. These scores are more appropriate to use if you want to make comparisons between students or aggregate scores, for example, to look at whether a whole class of students is progressing as expected.

Standard scores compare the student's performance to a large normative sample. By converting a student's raw score to a standard score, you can see how typical their performance is compared to other students that are similar to them. Age standardised scores allow you to compare their performance to same age peers. Year group standardised scores allow you to compare to students of the same academic year group. The average performance of the normative sample is converted to a standard score of 100 with a standard deviation of 15. This means that 68% of students would receive a standard score between 85 and 115, so this is usually considered the normal or average range of ability. As a rule of thumb, the following classifications are usually applied

- 69 and below: Low

- 70-84: Below average

- 85-115: Average

- 116-130: Above average

- 131 and above: High

- Two further types of scores are often provided - percentile ranks and age equivalent scores. While these are based on normed samples, for statistical reasons, these scores cannot be meaningfully combined, added or subtracted. This means that these scores are not a good measure to use to examine a whole class, or to track progress over time.

Percentile ranks tell you about a student's performance as a percentage of the population who can be expected to perform as well or less well than them. A student with average performance receives a percentile rank of 50 – this student performed as well or better than 50% of the normative sample. Care should be taken not to confuse percentiles with percentage accuracy.

Age equivalents indicate the age at which the average student achieves a given raw score. Age equivalents should be used with caution and should not be used to inform decisions about how or what a student should be taught. For example, if an 8-year-old achieves a reading age equivalent of a 16-year-old, that only means that they answered the same number of questions correctly. It does not mean that they have the knowledge to read, understand or enjoy books that are intended for 16-year-olds!

Confidence intervals and **standard errors of measurement** are sometimes presented in addition to some of the scores listed above. These are valuable reminders that any assessment score is just an estimate of a student's true ability and will be subject to some degree of measurement error.

Glossary

Age equivalents: norm-referenced tests compare a student's performance to those of a large sample of students. The age equivalents derived on these tests indicate the typical age at which the average student achieves a given raw score.

Confidence intervals and standard error of measurements: remind us that any test score is just an estimate of a student's true ability and will be subject to some degree of measurement error.

Criterion-referenced test: checks whether students have particular knowledge or skills which they are expected to have at a certain point. Does not give information about how performance compares to other students.

Decoding: using letter-sound rules to read a word.

Digraph: combinations of letters that represent a single phoneme. For example, the letters <sh> make a single sound. A digraph is a two letter grapheme.

Generalisation: using information learnt in one instance to guide your answer in a similar, but novel, situation, item or task.

Inferential reasoning: drawing a conclusion, or gaining a new piece of understanding, by thinking through and combining different pieces of information. In reading comprehension, questions requiring inferential reasoning can be contrasted to literal comprehension, in which the answer is explicitly stated in the text.

Irregular words: words that cannot be read using letter-sound rules alone (e.g., yacht).

Nonwords: made up or nonsense words.

Norm-referenced test: compares an individual student's performance to the performance of a larger sample. Gives information about whether their performance is average, or above/below average for their age or stage of education.

Percentile ranks: norm-referenced tests compare a student's performance to those of a large sample of students. The percentile ranks derived on these tests provided tell you what percentage of the normative sample performed at or below their level.

Raw scores: the total number of correct responses or score achieved on a test.

Regular words: words which can be read (or decoded) using letter-sound rules alone (e.g., cat).

Standard scores: norm-referenced tests compare a student's performance to those of a large sample of students. The standard scores derived on these tests provided tell you how typical an individual's performance is compared to other students.

References

Dockrell, J.E., Connelly, V., Walter, K., & Critten, S. (2014). Assessing children's writing products: the role of curriculum based measures. British Education Research Journal, 41(4), 575-595.

Daffern, T., & Critten, S. (2019). Student and teacher perspectives on spelling. Australian Journal of Language and Literacy, 42(1), 40-57.

Freeman, L., & Miller, A. (2001). Norm-referenced, criterion-referenced and dynamic assessment: What exactly is the point? Educational Psychology in Practice, 17(1), 3-16.

Ross, J.A. (2010). Effects of running records assessment on early literacy achievement. Journal of Educational Research, 97(4), 186-195.

PREPARED FOR THE EDUCATION HUB BY

Professor Julia Carroll

Julia is a professor of Child Development and Education at the Centre for Global Learning in Coventry University, UK. Julia is interested in understanding the many cognitive and non-cognitive factors that influence literacy development, particularly in children with difficulties.

Dr Helen Breadmore

Helen is an Associate Professor at the Centre for Global Learning in Coventry University, UK. Helen is interested in understanding the cognitive and linguistic processes that underpin reading and writing, and how to deliver best practice in literacy education.

Helen and Julia have worked on several projects to provide research-based guidance for teachers, including a review of the science of literacy development for the Education Endowment Foundation, UK.