

The purpose of assessment in schools

The New South Wales (NSW) Legislative Council Education Committee recently launched an inquiry into measurement and outcome-based funding in NSW schools. The MultiLit Research Unit was invited to make a submission to the committee regarding the types and purposes of literacy assessments.

The MultiLit Research Unit welcomes the opportunity to make this submission to the Education Committee's inquiry. We fully support the use of educationally sound and fit-for-purpose assessment to ensure all children learn to read.

All children should be reading proficiently by the time they finish primary school. A large number of children in NSW begin secondary school with levels of literacy that are too low for them to succeed at this level of education. According to the 2018 NAPLAN report, 5.5% of students in Year 7 in NSW schools did not achieve the very low national minimum standard. A further 12.7% only just achieved the standard. Together, these percentages represent more than 16,000 students beginning their secondary education as struggling readers.

The MultiLit Research Unit (MRU) does not take a position on the way funding allocations are made to schools. This level of policy advice, including the management of the complex array of incentives and disincentives created by outcomes-based funding, is beyond the MRU's remit. Our advice is limited to the most valid and appropriate assessments at various stages of children's reading development, and their fitness for purpose.

Purpose of assessment

Assessments may be conducted for one of several reasons:

- To screen students/early intervention
- To guide instructional decision-making
- To inform instructional grouping
- To gather evidence/diagnose disability/gain concessions
- To place students in programs/referral to specialist services
- To monitor progress/efficacy of an intervention/reporting
- To compare students against other students of same age/grade
- To evaluate educational systems/ranking

Information gathering takes place across different settings and circumstances:

- Individual
- Class



- Year
- School
- Region
- State
- Country

The purpose of the assessment should determine the type of assessment

There are numerous types of assessment for reading. It is essential to use assessments that are fit-for-purpose.

1 Curriculum-based assessment

Curriculum-based assessment (CBA) allows students to demonstrate their level of knowledge and skills along a continuum that describes a particular aspect of the curriculum. Curriculum-based assessment can be used at a student or class level to determine where to begin or revise a teaching sequence. They can be criterion-referenced, which means that benchmarks can be set based on expected learning outcomes after a certain amount of teaching and learning (as opposed to age-based norms).

CBAs are available both as commercial product or as a free resource, or they can be teacher-developed. CBAs are suitable for school and system performance and for

accountability reporting if the benchmarks are appropriately aligned with a common curriculum.

2 Curriculum-based measurement

Curriculum-based measurement (CBM) is a type of CBA and is used to monitor individual student progress on a regular basis. CBM can also be used for screening, referral for intervention, and for instructional decision-making.

Curriculum-based measures are typically short assessments that can be used repeatedly. CBM of reading is usually a measure of fluency (number of words read accurately per minute), which is highly correlated with overall reading development. CBMs are not suitable for school and system performance and accountability reporting.

3 Standardised tests

Standardised tests are designed to assess and allow comparisons between the achievement of different students. Standardised assessments are norm-referenced, meaning that they have been administered to representative samples of students and a profile created to provide information about how a student's performance on the test compares with their peers. Norms can be created for each item or question in the assessment, for subsections that

assess certain abilities, or for the whole assessment.

They have strict administration procedures and can be administered in groups or individually. They provide reliable information but cannot measure small changes in progress and should not be used frequently.

Standardised assessments are useful for research purposes but are not usually helpful for teachers making instructional decisions. If a child performs poorly on a standardised test, they will usually require further assessment to determine appropriate instructional response. Used with careful reference to statistical limitations, they are suitable for school and system performance reporting and accountability.

4 Diagnostic assessment

Diagnostic assessments are used to identify specific strengths and weaknesses in the reading abilities of individual students. They assess reading and language subskills in order to generate a diagnosis for a student's reading difficulty.

Diagnostic assessments are used (often also with reference to standardised norms) to identify specific difficulties in the reading profiles of individual students. Most diagnostic

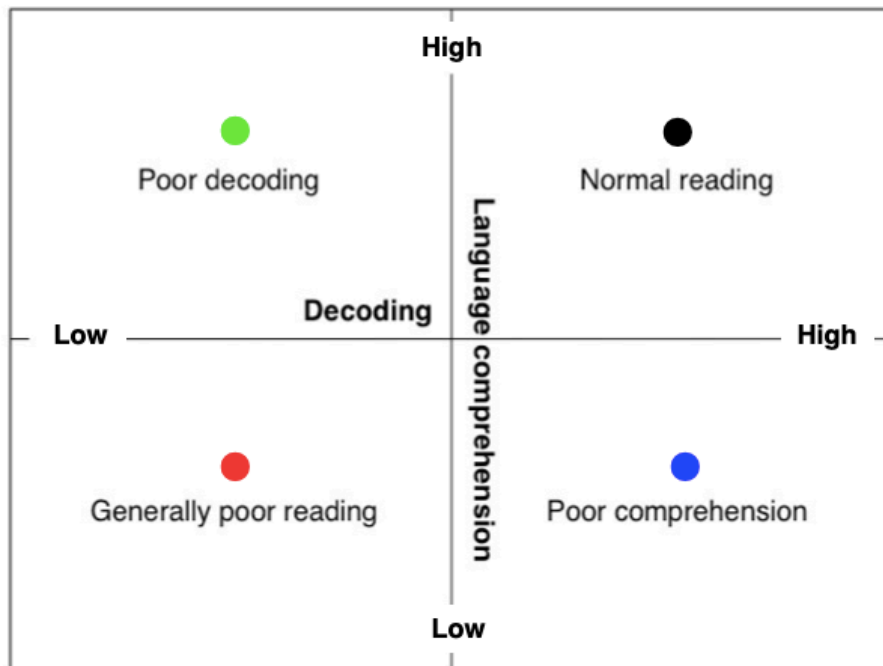


Figure 1. The Simple View of Reading

assessments are administered by allied education and health professionals, such as psychologists and speech pathologists.

Diagnostic assessments are necessarily time-consuming. They should usually be used only for students for whom more detailed information is required than can be provided by more general assessments of reading that are appropriate for their age and stage of learning (for example, from curriculum-based assessments and standardised tests). Diagnostic assessments should be used as and when a child demonstrates that they may be in need of specialised support and not administered at a set time in the school calendar. It is not appropriate to use diagnostic assessments for system performance monitoring and accountability purposes.

Assessment should be research-informed

Theoretical framework

The most strongly validated model of reading is the ‘Simple View of Reading’ (see Figure 1). It states that reading requires two areas of skill: decoding and language comprehension. If a child has difficulties with either of these components, their reading comprehension will be weak. This model provides a reference point for assessment.

Children start school with different levels of reading ability for a variety of reasons. As they progress through primary school, their teaching needs

will differ depending on their stage of reading development. Assessment should focus on what is important at each stage and what will make the difference to instruction.

In broad terms, in the first few years of school, teaching and assessment in reading should focus on developing letter-sound knowledge and decoding (phonics) and oral language as matters of priority. When children become proficient at decoding, other aspects of reading will dominate instruction, including a continuing focus on vocabulary and the more complex aspects of reading comprehension.

Response to Intervention

Assessments should inform a cascading ‘Response to Intervention’ (RTI) approach to ensure that children are given the type of assessment they require at the time that is required, in order for their learning to be effectively supported.

Response to Intervention typically has three levels or ‘tiers’ of teaching and assessment. Tier 1 is whole class teaching; Tier 2 is small group intervention for children who are having difficulty keeping up with their peers (usually the lowest 25%); and Tier 3 is specialist one-to-one instruction for children who still struggle to make progress after Tier 2 intervention.

A whole school RTI approach is efficient and effective: it supports early intervention, enables school management to measure the success of

school programs and make adjustments when required, and minimises the likelihood that a child will ‘fall through the cracks’.

Appropriate assessment depends on the stage of reading development

The following guideline represents the suggested types of reading assessments in a Response to Intervention model, and some examples of assessments that might be used; not all of them are necessary (see Table 1).

Assessments currently being used in NSW schools

Schools in NSW currently use a variety of reading assessments, some of which have been specifically developed for use in NSW schools (e.g., Best Start and PLAN2), some of which are part of the national assessment program (NAPLAN), and some of which are chosen at the discretion of the teacher or school. To what extent schools are using research-informed assessments with or without a Response to Intervention approach is not known.

The assessments currently used for school and system performance monitoring are the [NAPLAN tests](#). NAPLAN is a set of standardised tests with Australian student population-based norms (scaled scores) and which also has a criterion-referenced component (achievement bands). While the achievement bands are aligned to the Australian curriculum it is not a curriculum-based assessment per se.

The reading section of NAPLAN is a general comprehension measure. If a student obtains a low score, the test does not provide any information about the particular aspects of reading with which

Table 1. Assessments currently used in NSW schools

Stage	Tier	Example of assessments
Early Stage 1 & Stage 1	Tier 1 Curriculum-based phonics and oral language assessments; curriculum-based fluency measures	<ul style="list-style-type: none"> • Educheck (phonics) • Sutherland Phonological Awareness Test – R (phonological awareness and phonics) • Year 1 Phonics Screening Check (phonics) • Wheldall Assessment of Reading Lists (word identification fluency) • CELF-5 Screener (oral language) • CUBED Narrative Language Measures (oral language) • Dynamic Indicators of Basic Early Literacy Skills (phonics, vocabulary, fluency, comprehension)
	Tier 2 Curriculum-based phonics and oral language assessments; curriculum-based fluency measures	<ul style="list-style-type: none"> • Castles and Coltheart 2 (word reading/decoding) • Wheldall Assessment of Reading Lists (word identification fluency)
	Tier 3 Curriculum-based and diagnostic assessments	<ul style="list-style-type: none"> • Peabody Picture Vocabulary Test-4 (receptive vocabulary) • CELF-5 (oral language) • Castles & Coltheart 2 (word reading/decoding)
Stages 2 & 3	Tier 1 Curriculum-based and standardised reading comprehension assessments	<ul style="list-style-type: none"> • Wheldall Assessment of Reading Passages (oral reading fluency) • Progressive Achievement Tests in Reading (vocabulary and comprehension)
	Tier 2 Curriculum-based and standardised reading comprehension assessments	<ul style="list-style-type: none"> • Castles and Coltheart 2 (word reading/decoding) • Wheldall Assessment of Reading Passages (oral reading fluency) • Tests Of Reading Comprehension (TORCH) (reading comprehension) • York Assessment of Reading for Comprehension – Passage Reading (reading accuracy, rate, comprehension) • Neale Analysis of Reading Ability (reading accuracy, rate, comprehension) • Dynamic Indicators of Basic Early Literacy Skills (phonics, vocabulary, fluency, comprehension)
	Tier 3 Curriculum-based and diagnostic assessments	<ul style="list-style-type: none"> • Martin and Pratt Non-Word Reading Test (phonics) • Wechsler Individual Achievement Test-III (oral language, listening comprehension, decoding, word reading, reading comprehension) • Tests of Reading Comprehension (TORCH) (reading comprehension) • Waddington Diagnostic Standard and Advanced Reading and Spelling Test (Picture vocabulary knowledge, speech irregularities and potential word reversals, specific vowel differences, two letter consonant blending knowledge, ability to verify/discern beginning/endings of words, graphophonic and syntactic/semantic based skills, conceptual understanding of specific words, correct use of indirect picture cues/ comprehension of complex sentences)

The purpose of assessment in schools



they are having difficulty (decoding or language comprehension, or both) and therefore can only be used as an indicator of a student's reading ability that may need further investigation.

This is not a flaw in the assessment as it was not designed for individual diagnostic purposes. The NAPLAN assessment does have some serious deficiencies – particularly the poor item discrimination at the lower end of the range, which means the National Minimum Standard has poor reliability – but the standardised aspect of the assessment is not the problem. This form of assessment is the most appropriate one for large scale performance monitoring. Schools should be regularly using other assessments of the type described above for instructional purposes.

The first NAPLAN assessment takes place in Year 3, after children have had three years and one term of schooling. If a child is struggling with reading at this stage, it is difficult to remediate effectively. Ideally, NAPLAN would not be the first time that any child is identified as struggling with reading.

However, there are still a large number of children who perform poorly on NAPLAN, which suggests that for many of them, their reading difficulties have not been previously identified, they have not received effective intervention, or both.

For this reason, we recommend that an earlier systemic assessment be introduced to reduce the number of children who reach Year 3 without their reading difficulties being addressed. An expert panel appointed to advise the federal education minister on a Year 1 literacy assessment in 2017 recommended the adoption of the UK government's Year 1 Phonics Screening Check – a curriculum-based assessment that has a criterion-referenced expected standard of achievement¹. It meets the requirements for an assessment for this purpose. It has been a systemic assessment in South Australia since 2018, and a trial of the Year 1 Phonics Check has been announced in NSW in 2020. We welcome this decision and hope that the check will be adopted state-wide following the trial.

Conclusions

This submission is intended to give an overview of the purposes of assessment and the most appropriate assessments for those purposes. Assessment is only useful insofar as it provides reliable and accurate information that can be used to guide decisions on instruction or on policy, and lead to improvements in teaching and learning. It is therefore essential to make assessment decisions based on sound measurement principles.

MRU would like to thank Alison McMurtrie for her contribution to this document. MultiLit Pty Ltd is a developer of reading programs, resources, and assessments that reflect scientific best practice on reading instruction and intervention. It was an initiative of Macquarie University. Its work is guided by the MultiLit Research Unit which consists of six reading researchers with PhD level qualifications, led by Emeritus Professor Kevin Wheldall AM.

¹ The panel was chaired by a member of the MultiLit Research Unit – Dr Jennifer Buckingham