

More WARs: the development of the WARL and the WARN

How do teachers know how well their students are learning what they are being taught? **Kevin Wheldall and Robyn Wheldall**, of the MultiLit Research Unit and Macquarie University, describe the development of two very practical tests that teachers can use to monitor students' progress during the early stages of learning to read.

The MultiLit Research Unit has developed a series of assessment tools – curriculum-based measures (CBM) – that can be used to monitor the ongoing progress as students learn to read. In a previous issue of the LDA Bulletin, we reported the development of the *Wheldall Assessment of Reading Passages* or WARP, which can be used to assess the fluency with which students read passages of text (Wheldall, K., & Wheldall, R., 2020). The WARP is suitable for use with students who are reading at the Year 2 to Year 5 level (Wheldall & Madelaine, 2000; 2006). In this current article, we describe the development by the MultiLit Research Unit of two other curriculum-based

measures of reading fluency that are suitable for use with younger children who are performing at Year 1 and 2 levels: the *Wheldall Assessment of Reading Lists*, or WARL (Wheldall et al., 2015), and the *Wheldall Assessment of Reading Nonwords*, or WARN (Wheldall et al., 2021, in press).

It is very important to have CBMs that can track progress across the first two years of schooling while students are (ideally) learning to read via explicit phonics instruction, and to have an efficient way of identifying students who are not making typical progress in the early stages of learning to read. By administering a test that identifies struggling students effectively, as early in the process as possible, teachers may be able to address the needs of struggling students in a timely manner and also to monitor their progress. This will result in fewer students being left to struggle for longer than necessary (Bell et al., 2020).

There are relatively few tests that measure general reading progress satisfactorily in the early years and far fewer still that allow monitoring on a regular basis. The two CBM assessment tools to be discussed here focus on the reading of single words (the WARL), and the reading of nonwords (the WARN).

To be of any practical use, any test or measure must be both reliable and valid. The authors of the test must be able to provide empirical evidence for the *validity* and *reliability* of their test. By validity, we mean the degree to which a test measures what it is supposed to measure. One of the most common ways of verifying if a new test is valid is by correlating the scores on the new test with scores on older tests that have already been established as



valid indicators of reading performance (*criterion validity*). By reliability, we mean that the instrument must be capable of delivering the same result consistently. The test should give the same (or a very similar) result when it is given to the same child on separate occasions close together in time. If Mark scores 43 on the test on Monday, and assuming that he has not been practising in between, then he should get a very similar score to 43 on, say, Wednesday, if the test is reliable. We call this *test-retest reliability*. Similarly, if the test has two different forms, say Form A and Form B, then they should provide very similar results. We call this *parallel forms* reliability. The most common measure of reliability is the correlation coefficient between the scores of the test on the two occasions it is given, or between the two forms of the test when they are given to a group of children.

This article will describe the construction of the WARL and the WARN and provide data on reliability and validity for both tests. This article also provides references to research we have carried out for the purposes of providing benchmark guidelines for the WARL and WARN. These benchmarks are guides based on a small but reasonably representative sample of students. Students who score below the score designating the 25th percentile

(bottom quartile) may be considered to be 'struggling' or low-progress readers and in need of reading intervention support. The 40th percentile scores provide minimum goals for students to achieve before exiting an intervention, in that scores within the 40th to 60th percentile range may be considered to be within the average range for literacy performance for that point in the school year. We hope that these benchmarks will provide rough approximations to guide instructional decision-making. It should be noted, however, that these are not 'norms' in the strict sense of being based on large representative samples of students.

Another brick in the WARL

We would like to acknowledge, at the outset, the major contribution of Dr Meree Reynolds in the development of this measure as part of her doctoral studies.

The Wheldall Assessment of Reading Lists (WARL) consists of fifteen word lists. To construct the lists of words for the WARL, we started with a database of the 200 most common high frequency single words found in children's storybooks and reading schemes read by five- to seven-year-old children (Stuart et al., 2003). These 200 words were arranged into 20 groups of 10 words, with the words with the highest frequency being used in the first group and so on. Five words were randomly selected from each of these 20 groups and presented on a stimulus sheet as a 100-word reading task. This procedure was repeated 15

times to produce 15 alternative forms of the curriculum-based measure, each comprising 100 words.

The fifteen 100-word lists created were administered to a sample of 112 Year 1 students, who read each list for one minute each. Descriptive statistics for the 15 WARL lists (see Reynolds et al., 2009) showed that the means and standard deviations of the word list measures were relatively similar. Two of the word lists were subsequently excluded by a process in which consideration was given to both outliers and intercorrelations.

Following the procedure used in developing the WARP (see Wheldall & Wheldall, 2020), a decision was made to select three word lists from the remaining 13 lists, to be designated as the Initial Assessment Reading Lists. They were selected on the basis that they had the most similar means and standard deviations for words read correctly per minute. In addition, they correlated very highly with each other. The set of three Initial Assessment Word Lists of the WARL was deemed to be appropriate for screening procedures, for placement of students at appropriate levels of support, for pre- and post-testing in research studies, and for program evaluation. The mean of performance on the three lists is taken as the most reliable index, expressed in terms of words read correctly per minute.

The 10 word lists that remained were designated for monitoring progress during an intervention. The lists were very similar to one another in relation to

their means and standard deviations. They also correlated highly with each other and with the mean score of the three Initial Assessment Lists. We suggest that if two WARL lists are administered fortnightly and averaged, the data is likely to be more reliable, smoother and more even in increments, enabling easier interpretation. We have produced a designated order in which the Progress Monitoring Lists are used. When used in this order, the mean of each two successive progress tests is very similar.

Reliability and validity data for the WARL is summarised in Table 1 below.

Benchmark values for the WARL were subsequently calculated (Reynolds et al., 2011), for the average and bottom quartile scores for students at the beginning and middle of Years 1 and 2, as a guide for classroom teachers regarding typical progress.

Be WARNed

Measures of phonological recoding (nonword reading) and measures of reading fluency for students in the first two years of schooling are uncommon. (See Colenbrander et al., 2011, for a review of nonword tests.) The *Martin and Pratt Nonword Reading Test* (Martin & Pratt, 2001) measures nonword reading but is not timed and offers only two forms. The *Test of Word Reading Efficiency 2 (TOWRE-2)* (Torgeson et al., 2012) includes nonword reading and is timed but, again, has only two forms available. The *Year 1 Phonics Screening Check*, introduced by the UK Department of Education and now used in several sites in Australia (Department of Education, Skills and Employment, 2020) is a one-off test given at the end of Year 1 that includes a measure of nonword reading but is, again, not timed.

The *Wheldall Assessment of Reading Nonwords* (or WARN) is a new curriculum-based measure of nonword reading developed by the MultiLit Research Unit (Wheldall et al., 2021, in press). The measure is intended as a quick and simple test to measure progress in learning phonics decoding skills (phonological recoding) during the early stages of reading skill development, and to identify young struggling readers. The advantage of the WARN over existing measures of phonological recoding is that it comprises multiple parallel forms, thereby allowing for continual monitoring of individuals over time.

The WARN consists of 13 lists of 50 nonwords. Three of the lists are used

Participants: N = 122 Year 1 students (Reynolds et al., 2009)

Parallel form reliability	Intercorrelations amongst 15 individual WARL lists	WARL – all list intercorrelations: .80 to .97 (most coefficients over .90)
---------------------------	--	--

Participants: N = 335, Year 1 and Year 2 students. Assessed in February/March and again in August (Reynolds et al., 2011)

Parallel form reliability	Three Initial Assessment Lists on both testing occasions	WARL Initial Assessment Lists intercorrelations: .93 to .96
Test-retest reliability	Three Initial Assessment Lists, February/March and re-tested in August	List A test-retest: .82 List B test-retest: .84 List C test-retest: .86 Average t3 lists test-retest: .86
Criterion validity	Average 3 Initial Assessment Lists; Martin and Pratt NW Reading Test; South Australia Spelling Test; Sutherland Phonological Awareness Test-Revised (SPAT-R)	WARL and Martin & Pratt NW Reading Test: .75 WARL and Burt Word Reading Test: .87 WARL and South Australia Spelling Test: .83 WARL and SPAT-R: .83 WARL and WARP: .91

Table 1. Technical data (reliability and validity) for the WARL. All correlations significant at $p < .001$.

as the Initial Assessment Lists, and the remaining ten lists form five sets of two Progress Monitoring Lists, to be used fortnightly for the purpose of tracking progress. The Initial Assessment Lists can be used for screening or as a post-test measure following an intervention, either after two school terms or at other intervals.

Students read from each list for 30 seconds to determine the number of nonwords read accurately within that timeframe, and their performance over three lists (Initial Assessment Lists) or two lists (Progress Monitoring Lists) is averaged.

The WARN offers content validity, as the test stimuli align closely with the content sequence of *InitialLit Foundation (InitialLit-F)*, an instructional program which adheres to best practice according to the available theory and research (MultiLit, 2017). Nonword stimuli on the WARN were constructed using phonemes taught in the *InitialLit-F* program. The words in each list follow the sequence of the phonemes in the program, which in turn was based on the principles outlined by Carnine et al. (2006).

The *InitialLit-F* instructional program (MultiLit, 2017), which is targeted towards beginning readers, comprises 11 succeeding levels (known as 'sets') of instruction in letter-sound correspondences as part of a systematic synthetic phonics program. For the

purpose of constructing the WARN, Sets 1 and 2 were combined to form 10 'sets' in total. Ten nonwords were generated from each of the reduced sequence of sets, using the letter-sound correspondences taught at each successive set. The nonwords were three or four phonemes in length (CVC, CCVC or CVCC; C = consonant, V = vowel), and included digraphs (for example, fim, juck, nump, swong).

Each WARN list was created by randomly selecting five nonwords from the 10 nonwords constructed at each set, yielding a list of 50 nonwords presented on a stimulus sheet. This process of randomly selecting five words from 10 alternatives from each set was repeated 15 times to generate 15 lists, each comprising 50 nonwords.

All lists were administered to a sample of Foundation (first year of schooling) and Year 1 students. Means and standard deviations for each measure were calculated and all measures were inter-correlated. As expected, all 15 nonword lists produced very similar means and standard deviations and were highly inter-correlated ($r = .92-.96, p < .001$).

From these 15 lists, the most similar 13 lists were chosen and allocated to one set of three lists and five sets of two lists; the former to serve as the Initial Assessment Lists and the latter to serve as the Progress Monitoring Lists. The averages of these six sets were analysed

to confirm that they were highly inter-correlated ($r = .97-.98, p < .001$).

Reliability and validity data for the WARN is summarised in Table 2.

Benchmark values for the WARN were calculated for the average and bottom quartile scores for students in the first and second years of schooling, as a guide for classroom teachers regarding typical progress (Wheldall et al., 2021, in press).

Conclusion

Curriculum based measurement (CBM) is a quick, reliable, valid and cost-effective method of tracking progress in reading. It provides valuable information which enables educators to monitor progress regularly and to make appropriate instructional decisions in order to maximize the reading progress of their students. The series of CBM instruments we have developed (collectively known as the WARs) provide an effective Australian solution to progress monitoring of reading.

But what of the future? A problem upon which we are still working is the development of yet another WAR, the *Wheldall Assessment of Reading Comprehension* or WARC. This is proving more difficult but we continue to experiment with a maze procedure, whereby students need to select the seventh words from a 200 word passage from a list of four plausible alternatives. Watch this space!

Recommended reading on curriculum based measurement

Fuchs, L.S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review*, 33, 188-192.

Hosp, M. K., Hosp, J.L., & Howell, K.W. (2016). *The ABCs of CBM: A practical guide to curriculum-based measurement*. The Guilford Press.

Madelaine, A., & Wheldall, K. (1999). Curriculum-based measurement of reading: A critical review. *International Journal of Disability, Development and Education*, 46, 71-85.

Madelaine, A., & Wheldall, K. (2004). Curriculum-based measurement of reading: Recent advances. *International Journal of Disability, Development and Education*, 51, 57-82.

Tindal, G. (2013). *Curriculum-based measurement: A brief history of nearly everything from the 1970s to the present*. ISRN Education, 2013, 1–29. doi:10.1155/2013/958530

Study 1: Initial test development Participants: N = 163. Two schools similar to national average NAPLAN* Year 3. 85 Foundation** (First year of schooling) and 78 Year 1 students		
	Tests used	Correlation coefficients
Parallel forms reliability	WARN Initial Assessment set of lists and 5 sets of Progress Monitoring lists	WARN intercorrelations .97 to .98
Criterion validity	WARN Initial Assessment set of Lists and 5 sets of Progress Monitoring lists; Martin & Pratt; WARL	WARN lists and Martin & Pratt: .85 to .86 WARN lists and WARL: .91 to .92
Discrimination	WARN Initial Assessment set of lists, compared for Foundation and Year 1	Scores doubled from first to second year of schooling, showing good discrimination
Study 1: Initial test development Participants: N = 163. Two schools similar to national average NAPLAN* Year 3. 85 Foundation** (First year of schooling) and 78 Year 1 students		
Test-retest reliability	Three Initial Assessment Lists on both testing occasions	WARN test-retest .89
Criterion validity	WARN Initial Assessment set of lists; Martin & Pratt; WARL	WARN and Martin & Pratt: .90 WARN and WARL: .89

* NAPLAN: National Assessment Program – Literacy and Numeracy

** Foundation: first year of formal schooling

Table 2. Technical data for the WARN. All correlations significant at $p < .001$

References

Bell, N., Farrell-Whelan, M., & Wheldall, K. (2020). Use of early word reading fluency measures to predict pass rates on the Phonics Screening Check. *The Australian Journal of Education*, 64(2), 161-176. <https://doi.org/10.1177%2F0004944120931146>

Carnine, D.W., Silbert, J., Kame'enui, E.J., Tarver, S.G., & Jongjohann, K. (2006). *Teaching struggling and at-risk readers: A direct instruction approach*. Merrill Prentis Hall.

Colenbrander, D., Nickels, L., & Kohnen, S. (2011). Nonword reading tests: A review of the available resources. *Australasian Journal of Special Education*, 35(2), 137-172.

Department of Education, Skills and Employment. (2020). <https://www.dese.gov.au/australian-curriculum/year-1-phonics-check>. Retrieved 30 December 2020.

Martin, F., & Pratt, C. (2001). *Nonword Reading Test*. ACER Press.

MultiLit (2017). *InitiaLit-F: Whole class instruction in literacy*. MultiLit Pty Ltd.

Reynolds, M., Wheldall, K., & Madelaine, A. (2009). Building the WARL: The development of the Wheldall Assessment of Reading Lists, a curriculum-based measure designed to identify young struggling readers and monitor their progress. *Australian Journal of Learning Difficulties*, 14, 89-111.

Reynolds, M., Wheldall, K., & Madelaine, A. (2011). Early identification of young struggling readers: Preliminary benchmarks for intervention for students in years one and two in schools in New South Wales. *Australian Journal of Learning Difficulties*, 16, 127-143.

Stuart, M., Dixon, M., Masterton, J., & Gray, B. (2003). Children's early reading vocabulary: Description and word frequency lists. *British Journal of Educational Psychology*, 73(4), 585-598.

Torgeson, J.K., Wagner, R.K., & Rashotte, C.A. (2012). *Test of Word Reading Efficiency 2 (TOWRE 2)*. Pearson.

Wheldall, K., & Madelaine, A. (2000). A curriculum-based passage reading test for monitoring the performance of low-progress readers: The development of the WARP. *International Journal of Disability, Development and Education*, 47, 371-382.

Wheldall, K., & Madelaine, A. (2006). The development of a passage reading test for the frequent monitoring of performance of low-progress readers.

Australasian Journal of Special Education, 30, 72-85.

Wheldall, K., Reynolds, M., & Madelaine, A. (2015). *The Wheldall Assessment of Reading Lists (WARL)*. MultiLit Pty Ltd.

Wheldall, K., Reynolds, M., Madelaine, A., & Bell, N. (2021, in press). *The Wheldall Assessment of Reading Nonwords (WARN) Manual*. MultiLit Pty Ltd.

Wheldall, K., & Wheldall, R. (2020). Mentioning the WARs: Let's do the timed WARP again. *LDA Bulletin*, 52(1), 17-19.

Emeritus Professor Kevin Wheldall, AM, BA, PhD, C.Psychol, MAPS, FASSA, FBPsS, FCollP, FIARLD, FCEDP, served as Professor and Director of Macquarie University Special Education Centre (MUSEC) for over twenty years prior to his retirement in 2011. He is Chairman of MultiLit Pty Ltd and Director of the MultiLit Research Unit and is the author of over three hundred academic books, chapters, and journal articles. In 1995, he established the MultiLit (Making Up Lost Time In Literacy) Initiative, to research and develop intensive literacy interventions. He is a Fellow of the Academy of Social Sciences in Australia and in 2011 was made a Member (AM) in the Order of Australia.

Dr Robyn Wheldall (formerly Beaman), BA, PhD., was a Research Fellow at Macquarie University until her retirement in 2011 and now continues as an Honorary Fellow. She is a founding director of the University spin-off company MultiLit Pty Ltd and is the Deputy Director of the MultiLit Research Unit. She jointly authored 'An Evaluation of MultiLit' (2000) (commissioned by the Commonwealth Government) and has published numerous articles in peer reviewed journals. Robyn has extensive experience in the establishment and implementation of intensive literacy programs in community settings. In 2005 she was awarded a Macquarie University Community Outreach Award for her MultiLit work.

Disclosure

Kevin and Robyn Wheldall are directors of MultiLit Pty Ltd, in which they have a financial interest. They receive a benefit from the activities of the company and the sale of its programs and products, including the measures that are the subject of this article.