

Mentioning the WARs: Let's do the timed WARP again

In this article **Kevin and Robyn Wheldall** present an alternative curriculum-based measurement that assesses reading fluency. They present interesting behind-the-scenes information about the development of the Wheldall Assessment of Reading Passages, or WARP, and discuss the use of the test for the ongoing monitoring of reading progress. Further developments in the Wheldall WARfare suite of assessment tools will appear in a subsequent LDA Bulletin!

The assessment of reading ability has a long history in educational psychology and special education. Burt, Schonell, Vernon, Neale, to name but a few, all offered what were known as 'reading tests', to assess the progress of children's reading ability, typically expressed as a reading age (akin to the more general concept of mental age). Children whose performance was substantially behind that of their peers could thereby be identified and offered 'remedial' assistance. One of the things

that these tests had in common was that they were quite time-consuming. Even using a very simple test like the Burt took a long time to assess a whole class of children. If only a quicker and simpler measure were available... Another problem was that these standardised reading tests could (or should) only be used infrequently; say, every six or twelve months because of practice effects. Some of these tests offered parallel forms but this barely scratched the surface of the problem. Most reading tests are also insensitive to small changes in reading progress. Educators need to monitor the reading progress of low-progress readers on a very regular basis, in order to make instructional decisions well before the conclusion of a program or the end of a school year.

Educators need to monitor the reading progress of low-progress readers on a very regular basis, in order to make instructional decisions ...

Curriculum-based measurement (CBM) is a method of assessing growth in basic skill areas. One skill area where this has been widely employed is that of reading. Several curriculum-based measures of reading exist but perhaps the most widely used is oral reading fluency (ORF). ORF is measured by a passage reading test, which requires students to read aloud from a passage of text for one minute, to determine the number of words read correctly per minute. Research on CBM of reading dates back to the early 1980s and continues to the present day. As such, CBM of reading has a large and very sound research base. Many studies have provided evidence of the reliability and validity of CBM of reading. ORF has been found to be a valid indicator



of general reading ability including reading comprehension.

An essential feature of this assessment method is that test materials are drawn from the students' curriculum, originally taken directly from a basal reading series. By reading a passage of text, the whole skill of reading is measured, rather than component sub-skills. Research has also demonstrated that CBM of reading is an effective means of monitoring reading progress, particularly that of low-progress readers on, say, a weekly or fortnightly basis, using a set of curriculum-based passage reading tests. This information is then used to make instructional decisions such as increasing the intensity or frequency of instruction and is ideally suited for use within a Response to Intervention (RTI) model.

Too good to be true?

We first became acquainted with curriculum-based measurement (CBM) of reading in the early 90s, when we began to read the pioneering research of Stan Deno and his colleagues (Deno, 1992; Deno et al, 1982). Quite frankly, it all sounded too good to be true initially. Could it really be the case that one could assess reading progress accurately and reliably by asking a child to read from a passage of text for just one minute and then counting the number of words read correctly? We were dubious. To be convinced we had

to collect data of our own; we did and we were.

Could it really be the case that one could assess reading progress accurately and reliably by asking a child to read from a passage of text for just one minute and then counting the number of words read correctly?

Our first attempts involved using passages of grade level text from 'real books' from the curriculum, which were judged to be of about the same level of difficulty, as recommended originally by Deno. This proved to be quite challenging even when using readability formulae to estimate similar levels of text difficulty. Moreover, for our purposes, working with low-progress readers differing in age, we needed passages that were not necessarily grade related; passages that could be used across grades. It was subsequently determined that such passages need not be literally based in the curriculum, defined narrowly i.e. the actual books children were reading in class. Fuchs and Deno (1994) asked 'Must instructionally useful performance assessment be based in the curriculum?' and concluded that it did not. They interpreted the relevant curriculum as the broader concept of reading *per se* and that specially composed, novel passages could be used equally well.

Doing the timed WARP again

To this end, the first author (KW) wrote a series of 21 200-word passages of narrative text, each comprising a simple, short story. We checked and adjusted the draft passages based on the readability measures provided in Microsoft Word, to make them as similar as possible in terms of reading difficulty. But it soon became clear from our pilot studies that this was not sufficient. The only reliable way of developing parallel passages was to try them out on relevant samples of children (Wheldall & Madelaine, 1997). Dr Alison Madelaine was the major contributor to this enterprise, as part of her doctoral studies, and also compiled extensive reviews of the relevant literature (Madelaine & Wheldall, 1999; 2004). Literally hundreds, if not thousands, of students were assessed

on successive versions of what became known as the Wheldall Assessment of Reading Passages or WARP, over a period of several years, to establish its psychometric credibility and to provide performance benchmarks for successive school years. The published edition of the WARP comprises three Initial Assessment Passages and ten Progress Monitoring Passages.

What follows is a brief summary of the process by which the current WARP passages were selected and is fully described in Wheldall and Madelaine (2006). This version of the WARP derives from an analysis of a sample of 261 school students from Years 1 to 5 from the same school. As such, and while clearly not constituting a random sample of students in any sense, it comprised almost the total intake of students from Years 1 to 5 (the likely range of the test) from a school that had been shown to be closely representative of the population of school students in New South Wales over three successive years. This sample of students were all assessed by trained research assistants on all 21 of the 200-word passages.

The results, in terms of basic descriptive statistics and correlations for all 21 passages are provided in Wheldall and Madelaine (2006). In essence the results of preliminary analyses replicated all previous WARP studies in that all of the WARP passages were shown to intercorrelate very highly ($r=0.95+$), with very similar standard deviations. Mean numbers of words read correctly per minute for the 21 passages (i.e., the difficulty levels of the passages) varied, however. This was in spite of attempts to write all of the passages so as to be at the same level of difficulty and using readability measures. Consequently, the two easiest passages were discarded, as were the six most difficult passages, which were appreciably more difficult than the others. This left 13 passages of a very similar level of difficulty, as determined *empirically* by these results.

A decision was taken to select three passages, which were the three passages most similar to each other, and to deem that the mean score for this basic set of three Initial Assessment Passages be used as a set for 'one off' testing for screening and/or placement purposes, for termly assessments and reporting, for evaluation studies, etc. The three passages were very similar in terms of both mean and standard deviation for words read correctly and also intercorrelated very highly both with

each other ($r=0.97$) and mean passage score over the three passages (0.99).

The remaining ten passages from the 13 passages selected on the basis of their similarity to each other were chosen to yield a set of ten Progress Monitoring Passages. Following an initial assessment, these passages could be used weekly over the course of a typical ten-week term to monitor the progress of individual students. (A more reliable index of progress, reducing the error variance, may be obtained by calculating the running mean of these passages over the weeks or by taking the mean of two successive passages given every fortnight.) The ten passages were similar in terms of both mean and standard deviation for words read correctly, every passage mean being within four points of the mean for the three Initial Assessment Passages and the standard deviation varying by no more than three points from that for the average for the three Initial Assessment Passages. The ten passages also intercorrelated very highly with each other ($r = 0.95-0.98$) and with the mean passage score of the three Initial Assessment Passages ($r = 0.97-0.98$).

... the basic set of three Initial Assessment Passages be used as a set for 'one off' testing ... The remaining ten passages ... could be used weekly over the course of a typical ten-week term to monitor the progress of individual students.

Moreover, the passages showed good validity, confirming the results of our earlier studies. In a study comprising 146 low-progress readers, validity coefficients of 0.8 (range 0.78-0.80) were found between the WARP mean and the reading accuracy measure on the Neale and 0.52 on the Neale comprehension measure (Madelaine & Wheldall, 1998). A subsequent study sampled the full range of reading ability ($n=50$) and found higher correlations. The validity coefficients for the WARP and Neale Accuracy were again high at 0.87 (range for individual passages: 0.84-0.87); 0.71 (0.67 to 0.72 for individual passages) for Neale Comprehension; and 0.85 (range 0.83-0.85) for the Burt.

Given their similarity to each other and to the Initial Assessment Measure,

their use as parallel Progress Monitoring Passages would therefore appear to be warranted for successive use in monitoring reading progress, following a specific intervention, for example. The passages were deliberately ordered for use, so as to distribute the small differences between passages in such a way that they almost cancel each other out (when running means over two successive passages are calculated, for example). It is recommended that these data obtained be graphed to monitor continuing progress of individual students.

CBM is a quick, reliable, valid and cost-effective method of tracking progress in reading, providing valuable information which enables educators to monitor progress regularly and to make appropriate instructional decisions ...

We have developed other CBM measurements (collectively known as the WARs), as we develop and evaluate our own suite of reading programs. We will describe the other WARs in an upcoming LDA publication. For now, however, our experience is showing that CBM is a quick, reliable, valid and cost-effective method of tracking progress in reading, providing 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. Watch this space for the next time we mention the WARs!

References

Deno, S. L. (1992). The nature and development of curriculum-based measurement. *Preventing School Failure*, 36, 5-10.

Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children*, 49, 36-45.

Fuchs, L. S., & Deno, S. L. (1994). Must instructionally useful performance assessment be based in the curriculum? *Exceptional Children*, 61, 15-24.

Madelaine, A., & Wheldall, K. (1998). Towards a curriculum-based passage reading test for monitoring the performance of low-progress readers using standardised passages: A validity study. *Educational Psychology*, 18, 471-478.

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.

Wheldall, K., & Madelaine, A. (1997). Should we measure reading progress and if so how? Extrapolating the curriculum-based measurement model for monitoring low-progress readers. *Special Education Perspectives*, 6, 29-35.

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.

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 measure that is the subject of this article.