A Curriculum-Based Passage Reading Test for Monitoring the Performance of Low-Progress Readers: the development of the WARP

KEVIN WHELDALL & ALISON MADELAINE
Macquarie University Special Education Centre, Macquarie University, Sydney NSW 2109, Australia

ABSTRACT Results are presented from two studies involving 190 primary and high school (Year 7) students and 146 older low-progress readers in the late primary years. These studies contribute towards the development of five 200 word standardised passages which may be employed interchangeably to monitor the performance of low-progress readers towards achieving functional literacy. The passages are shown to possess both high parallel form reliability (0.94 to 0.96) and internal consistency (0.97 to 0.99), and also good criterion validity (0.78 to 0.80) with reading accuracy. Appropriate growth in performance over the primary years is also demonstrated. These findings suggest that an extrapolated model of curriculum-based measurement (CBM), which incorporates elements of more traditional reading assessments, would overcome some of the problems associated with CBM.

Introduction

Effective monitoring of the progress of low-progress readers is critical for effective remediation. If we are to remediate effectively with low-progress readers it is essential to know whether the individual child is progressing and, if so, at what rate the child is progressing. Empirical research strongly supports the use of such ongoing measurement and has found that it leads to measurable gains (see e.g., Fuchs, Deno, & Mirkin, 1984; Fuchs & Fuchs, 1986).

Deno, Mirkin, and Chiang (1982) suggested the Passage Reading Test (PRT) as an alternative, curriculum-based measure of reading progress to more traditional reading tests. The PRT requires students to read from any appropriate grade level basal reader for just one minute. The number of words read correctly in this period is the index of the student’s current level of reading performance. This simple measure of reading has been shown repeatedly to correlate highly with other measures of both reading accuracy and reading comprehension (Deno, Mirkin, & Chiang, 1982; Fuchs, Fuchs, & Maxwell, 1988). Moreover, such measures,
repeated say, twice weekly (as Deno, Mirkin, & Chiang recommend), may be used to track reading progress and, thereby, the efficacy of reading instruction provided.

An extensive review of the literature on the curriculum-based measurement of reading has recently been provided by Madelaine and Wheldall (1999) and, hence, only a brief overview will be provided here. Curriculum-based measurement (CBM) is a set of procedures for measuring student proficiency and indexing growth in the curriculum (Deno, 1987). This set of procedures was developed in response to the problems associated with the use of standardised, norm-referenced tests. Traditional tests are said to be problematic for a variety of reasons, including lack of content validity, lack of sensitivity to small changes in progress, absence of sufficient parallel forms to monitor progress on an ongoing basis, cost, and the time needed to administer them (Fuchs, Fuchs, & Hamlett, 1989b; Fuchs, Fuchs, & Maxwell, 1988; Jenkins & Jewell, 1993; Marston & Magnusson, 1988; Rodden-Nord & Shinn, 1991; Shinn, 1989).

Curriculum-based measures have been developed in areas such as spelling, mathematics, social adjustment, written expression and, particularly, reading (Deno, 1985; Deno, Marston, & Mirkin, 1982; Deno, Mirkin, & Chiang, 1982; Fuchs, Fuchs, Hamlett, & Allinder, 1991; Marston & Magnusson, 1988; Shinn, 1989). They have been used for a variety of purposes including progress monitoring, screening, referral, and instructional decision-making (Deno, 1987; Deno & Fuchs, 1987; Fuchs & Deno, 1992; Mehrens & Clarizio, 1993).

Reading is more typically measured using published standardised reading tests (Fuchs, Fuchs, & Hamlett, 1989b; Fuchs, Fuchs, & Maxwell, 1988) or informal teacher-made tests or observations (Deno, 1985; Fuchs & Deno, 1991). Several authors have identified the difficulties associated with using standardised reading tests to assess reading proficiency as noted above (Fuchs, Fuchs, & Hamlett, 1989b; Jenkins & Jewell, 1993), while the technical characteristics of teacher-made tests are unknown, rendering their usefulness as assessment tools questionable at best. Consequently, CBM of reading is regarded as a better indicator of general reading ability and progress (Potter & Wamre, 1990) since empirical research has clearly demonstrated the effectiveness of CBM in instructional decision-making (Deno, Mirkin, & Wesson, 1984; Fuchs, Fuchs, & Hamlett, 1989a).

Much of the research on CBM with regard to reading has focused on determining the best measure of reading progress (Deno, 1985; Deno, Mirkin, & Chiang, 1982). The available research indicates that Oral Reading Fluency (ORF), combining both accuracy and rate, is an accurate measure of general reading ability including reading comprehension (Deno, Mirkin, & Chiang, 1982; Fuchs, Fuchs, & Maxwell, 1988; Jenkins & Jewell, 1993). It also addresses many of the disadvantages of standardised tests. ORF is typically measured by a PRT and is expressed as the number of words read correctly per minute (Hasbrouk & Tindal, 1992). As previously stated, high correlations have been found between reading aloud measures and reading comprehension (literal and inferential) measures on standardised reading tests (Deno, Mirkin, & Chiang, 1982; Fuchs, Fuchs, & Maxwell, 1988).

If performance on curriculum-based passage reading tests is to be used to monitor progress toward a long term goal, that goal needs to be identified. We suggest
functional literacy as an appropriate minimum long term goal for low-progress readers. There are a number of definitions of functional literacy but, as several authors recognise, these are vague at best (Graff, 1995; Luke & Gilbert, 1993). If functional literacy is to represent an outcome, long term goal, or minimum level of competence in reading, then it needs to be clearly defined in terms of a level of reading proficiency, possibly linked to a reading age. There appears to be little research attaching a reading level to functional literacy, but a study on adult literacy by Goyen (1977) estimated it to be at around 10 years 3 months or, more generally, the level typically achieved by most students by the end of Year 5.

More recent literature has looked at the idea that CBM may not have to be based in the actual curriculum per se to be effective, and that perhaps it is better if it is not (Fuchs & Deno, 1994). Reasons include variations in readability between basal reading schemes, variations in readability between books supposedly at the same level in the same reading scheme, variations in readability within books, familiarity with passages, and the fact that passages taken from books do not represent a complete story (and therefore comprehension may be affected).

We acknowledge it could be argued that a set of passages that do not come from the actual curriculum (i.e., the reading scheme(s) or other material used in the classroom) is not truly curriculum-based in the strict sense. This is a fairly narrow definition of CBM, however. Using a broader definition, anything that students read may be considered as “the reading curriculum.” With this in mind, and with a view to providing a means of tracking reading progress, particularly of low-progress readers, to a level that represents functional literacy, a set of generic passages began to be developed at Macquarie University Special Education Centre (MUSEC) by the first author in 1995 (Wheldall, 1996). In its original form, the Wheldall Assessment of Reading Passages (WARP) consisted of fourteen 200-word passages. These passages, each comprising an entire story, were specially written to preclude the possibility of students having encountered the passages before. The readability of these passages was calculated using the measures within Microsoft Word and they were found to be of roughly equal difficulty. The passages all had a Flesch Reading Ease of 90 to 100 and a Flesch-Kincaid Grade Level of between 5 and 6 (Microsoft Corporation, 1993–1995). For a sample passage, see Appendix A.

It was originally thought that it would be better (both more reliable and more valid), to require students to read the whole 200-word passage. Not only would scores be thereby based on a larger sample of words, but every student would read the same words and the same number of words (as all passages are 200 words in length), so comparing scores would be comparing like with like. When using a one minute sample only, there is the risk of differences in difficulty level within passages influencing scores. For example, if a student reads 50 words in a minute and another student reads 100, the second 50 words may not be at the same level of difficulty as the first 50 words. It is for this reason that we decided to determine two fluency measures for each passage: the number of words read correctly in the first minute (WPM) and the number of words read correctly per minute over the whole passage (WPP).
All 14 passages were administered to 21 primary aged low-progress readers attending the Making Up Lost Time in Literacy (MULTILIT) Program at MUSEC. On the basis of these data, five passages were selected for further study. The five passages chosen correlated highly with each other (0.88 to 0.96 for both WPM and WPP) and, equally importantly, also had means and standard deviations that were very similar to each other (Wheldall & Madelaine, 1997). These results were subsequently confirmed by a second small scale study in which the five passages chosen were administered to another group of 21 low-progress readers from Years 6 and 7 attending an intensive community-based literacy intervention program linked to the MULTILIT Program at MUSEC. Intercorrelations between the five passages ranged from 0.83 to 0.96 for both forms of scoring, with means and standard deviations again similar (Wheldall & Beaman, 2000).

The correlation between the two measures of oral reading fluency (WPM and WPP) ranged from 0.96 to 0.99 for the five passages across the two studies. These are extremely high correlations. Consequently, it appears that the small increase in reliability and validity gained by having the student read the entire passage is not worth the extra time this takes. The consistently high relationship between the number of words read correctly in the first minute and the number of words read correctly per minute over the whole passage, indicates that one is a very good indication of the other. If basically the same information that may take up to 7 minutes to collect can be obtained in one minute, it is clearly a waste of time to use the second fluency measure (WPP) at all. Therefore, having demonstrated this, it is intended that a measure of oral reading fluency gained by having students read passages for one minute only, will be used in the further development of the WARP and in practice. Although WPP data was collected in the studies reported below, data will only be presented in so far as it provides measures of internal consistency, that is, the correlation between WPM and WPP.

This article reports the results of two larger scale studies that aimed to provide further complimentary evidence of the comparability of these five WARP passages, to provide evidence of internal consistency of the same five passages, and to establish preliminary estimates of criterion validity by examining the relationship between scores on the WARP and scores on another measure of reading. Study One also aimed to provide preliminary evidence for appropriate growth in reading performance over year (grade) levels.

Study One

Method

Participants. A group of 12 final-year undergraduate student teachers undertaking a University course in learning difficulties and 26 postgraduate students completing a Postgraduate Diploma in Special Education each administered the five selected passages to five school students as part of an assignment. Data collection was carried out in primary and secondary schools in Sydney. Schools included those in the State, Catholic, and Independent school systems. Classroom teachers were asked to identify four “average readers” and one “low-progress reader.” The identification of
these children was based solely on teacher judgement. This resulted in a sample of 190 participants (including 38 low-progress readers) in Years 3 to 7. The number of participants in each year were as follows: 20 in Year 3, 55 in Year 4, 20 in Year 5, 65 in Year 6, and 30 in Year 7.

Instrument. The original experimental edition of the Wheldall Assessment of Reading Passages (WARP) (Wheldall, 1996) consisted of a series of fourteen 200 word-passages, as already described. The five (WARP) passages chosen as a result of the small scale studies, as discussed above, were used in this study.

Procedures. The student data collectors were trained in procedures (modelled on Shinn, 1989) for administering the WARP. Participants were required to read aloud a series of passages for one minute. For a detailed description of the procedures, see Appendix B. A numbered column on the tester’s copy of each passage was used to assist in calculating the number of words read in one minute. From this score, errors were subtracted to calculate WPM. All sessions were audiotaped for the purpose of checking interobserver agreement.

Reliability. Interobserver agreement was calculated on 20% of the data and included all 38 data collectors. This was calculated by dividing the number of agreements by the number of agreements plus the number of disagreements and ranged from 94% to 100% with a mean of 99%.

Results and Discussion

In order to confirm that the five passages used in this study were of similar difficulty, correlation coefficients were calculated between the passages. The alternate forms reliability reported above in the small scale studies was replicated with this larger sample, with coefficients ranging from 0.95 to 0.96 (p < .001). In addition, internal consistency was extremely high, with correlations between WPM and WPP of 0.99 for all five passages (p < .001). Mean scores and standard deviations further support the similarity of the passages, with means ranging from 124 to 130 (Table I). T-tests confirmed that differences between the means for Passages 1 and 2 and between Passages 1 and 3 (the largest differences) were not significant (t = −1.50, p = .13 and t = −1.44, p = .15 respectively).

Table II shows descriptive statistics separately for each of the years used in the study for average students only, averaged over the five WARP passages. (Data for low-progress readers by year (grade) has not been presented due to very small

<p>| Table I. Descriptive statistics for Study One (N = 190) |
|---|---|---|---|---|---|---|</p>
<table>
<thead>
<tr>
<th>WPM</th>
<th>Passage 1</th>
<th>Passage 2</th>
<th>Passage 3</th>
<th>Passage 4</th>
<th>Passage 5</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>124</td>
<td>130</td>
<td>130</td>
<td>127</td>
<td>125</td>
<td>127</td>
</tr>
<tr>
<td>SD</td>
<td>36.14</td>
<td>40.02</td>
<td>41.50</td>
<td>38.41</td>
<td>41.09</td>
<td>38.65</td>
</tr>
</tbody>
</table>
TABLE II. Descriptive statistics by grade for average readers (Study One)

<table>
<thead>
<tr>
<th>WARP</th>
<th>Year 3 (n = 16)</th>
<th>Year 4 (n = 44)</th>
<th>Year 5 (n = 16)</th>
<th>Year 6 (n = 52)</th>
<th>Year 7 (n = 24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>105</td>
<td>127</td>
<td>134</td>
<td>155</td>
<td>147</td>
</tr>
<tr>
<td>SD</td>
<td>27.87</td>
<td>28.01</td>
<td>38.62</td>
<td>25.06</td>
<td>22.30</td>
</tr>
</tbody>
</table>

sample sizes.) Analysis of the data by year (grade) level revealed an interesting pattern. From these data, it appears that there is a substantial increase in WPM from Year 3 to Year 4 (22 WPM), followed by a much smaller increase from Year 4 to Year 5 (7 WPM). Scores then increased markedly again from Year 5 to Year 6 (21 WPM) and then decreased slightly from Year 6 to Year 7 (8 WPM). The data are also presented in Figure 1, which provides preliminary evidence of developmental growth in reading across year (grade levels).

When interpreting these data, three points need to be considered. Firstly, there were only 16 participants in each of Years 3 and 5. These small sample sizes could be affecting the apparently large increase from Years 3 to 4 and the apparently small increase from Years 4 to 5. Secondly, a ceiling effect may be responsible for the decrease from Years 6 to 7. For the entire sample, the correlation between chronological age and WARP mean was 0.40. If the participants in Year 7 are removed from the sample, this correlation increases to 0.43, providing further support for the presence of a ceiling effect. Thirdly, it needs to be remembered that the participants at each year (grade) level are different students, that is, this is cross-sectional, not longitudinal data. In other words, they are not strictly comparable.

![Fig. 1. Average reading growth across years.](image)
As previously stated, the participants in this study were classified as either "average" or "low-progress" readers. The mean WARP score for the average readers (138 WPM; \( SD = 31.6 \)) was 55 WPM greater than that for the low-progress readers (83 WPM; \( SD = 32.7 \)), which was highly significant \( (t = 9.51, p < .001) \). In other words, WARP very effectively distinguishes low-progress readers from regular readers. From the year (grade) level data displayed in Table II, it appears that the typical low-progress reader reads the WARP at below Year 3 level.

**Study Two**

**Method**

**Participants.** Study Two comprised a clinical sample of 146 low-progress readers from Years 5 to 8. The participants came from several sources but all had been referred to MUSEC and were under the remit of the MULTILIT initiative.

**Instruments.** Two instruments were used in Study Two. The first, the Neale Analysis of Reading Ability – Revised (Neale, 1988), is a widely used test which measures and provides reading ages for reading accuracy and reading comprehension, based on the results of an Australian standardisation. A recent study by McKay (1996) has confirmed the Australian version of the Neale as an accurate and reliable measure of reading ability. A reading age was determined for both accuracy and comprehension.

The second instrument utilised the same five passages from the Wheldall Assessment of Reading Passages (WARP) (Wheldall, 1996). The two fluency measures that were taken in the previous studies (WPM and WPP) were also employed, but again, data for WPP will only be reported where it relates to internal consistency.

** Procedures.** The Neale was administered according to the instructions in the Manual. The WARP was administered according to the procedures outlined in Appendix B, with one modification. It was noted that some students were taking a very long time to read the passages (more than 7 minutes per passage in some cases), which made the testing session extremely long and possibly distressing to the student. Where a student took more than 5 minutes to read the first two passages, they were only required to read the remaining three passages for one minute each. (This resulted in a slightly smaller sample size \( (N = 122) \) for some analyses involving the WPP measure.) Trained and experienced research assistants were employed to collect the data. A high level of reliability between testers had previously been established.

**Results and Discussion**

As in all of the other studies reported above, alternate forms reliability was found to be extremely high, with intercorrelations between the passages ranging from 0.91 to 0.93. Similarly, mean scores for the five passages (ranging from 75 to 82) and standard deviations were very similar (Table III). The difference between the means
for Passages 2 and 3, the largest difference between the five passages, was not significant ($t = 1.46$, $p = .15$). Measures of internal consistency were also high, with WPM and WPP correlating at 0.97 to 0.98.

Study Two also provided some criterion-related validity data, that is, a measure of how well the WARP correlates with another established reading test; in this case, the Neale. The participants in this study had a mean reading accuracy age of 101 months ($SD = 16.98$) and a mean reading comprehension age of 103 months ($SD = 17.57$). The average chronological age of the participants was 144 months ($SD = 11.74$) and, therefore, on average, the participants were reading at more than three years below their chronological age. The WARP mean correlated highly with the reading accuracy measure on the Neale, at 0.80 ($p < .001$), with correlations with individual passages ranging from 0.78 to 0.80. The correlation between Neale comprehension and the WARP mean was lower than had been expected at 0.52 ($p < .01$), with correlations with individual passages ranging from 0.49 to 0.55.

**Discussion and Conclusions**

If the WARP is going to be used to monitor progress frequently, the passages must be of equal (or as near equal as possible) difficulty, so that any gains in ORF reflect reading progress and not differences in passage difficulty. (Minor fluctuations in passage difficulty may be, and in practice are, typically overcome by the use of running means.) Alternate forms reliability has been consistently demonstrated in the above studies, with intercorrelations between the five passages ranging from 0.94 to 0.98 for both fluency measures across Studies One and Two. In addition to looking at the correlations between passages, mean scores and standard deviations also indicated great similarity across passages in terms of difficulty level. The high level of interobserver agreement obtained (almost 100% in most cases) reinforces the simplicity of employing this measure of reading progress in practice.

The correlations between WPM and WPP which we have used as a measure of internal consistency for the passages, have been consistently high (0.96 to 0.99) across all studies reported. As has been previously discussed, there is thus little point in requiring students to read the whole passage when words read correctly in one minute only provides almost identical results.

If this assessment is going to be used for tracking progress, standards need to be determined for either years (grades) or ages. The year (grade) level data supplied above provide tentative (only) indications of year (grade) norms. The participants in Study One, however, were not representative of the full range of reading ability as no “good” readers were included (only “average” and “low-progress” readers). In
order to establish norms for the WARP, a larger and more representative sample would be required, and this is currently being undertaken.

Similarly, the participants in Study Two are not representative of the full range of reading ability, as this was a clinical sample comprising only low-progress readers. This may account for the criterion validity coefficients being lower than expected, particularly between the WARP and Neale Comprehension. The correlation between Neale Accuracy and Neale Comprehension for this sample was only 0.65, for example, whereas for the Neale standardisation sample, the correlation between these two components was 0.90 (Neale, 1988). It is likely, then, that collecting similar data on a more representative sample of regular readers would yield higher criterion validity coefficients. Madelaine and Wheldall (1998) have demonstrated that this is, in fact, the case.

When considering what constitutes a reliable measure of reading progress, the level of difficulty of the measure must also be considered, particularly as this measure will be used to monitor progress towards a long term goal. A study by Mirkin and Deno in 1979 (cited in Fuchs & Deno, 1992) measured student performance using text at three difficulty levels. They found that material at frustration level was less sensitive to changes in progress than material at instructional or independent levels (Fuchs & Deno, 1992). This finding has implications for determining the minimum level in an optimal reading range associated with a particular measure; that is, the lowest reading level for which a particular measure is appropriate. Fuchs and Deno’s criterion for determining frustration level text is reading at a rate of fewer than 30 words read correctly per minute. Consequently, a passage on which a student reads fewer than 30 words per minute is unlikely to be a sensitive tool for measuring small changes in reading proficiency. (It may, however, still be appropriate for pre- and post-testing.) It should also be borne in mind that students may become distressed when required to read material at frustration level on an ongoing, long term basis.

It has been suggested that the range of years to which the WARP has been administered, that is, Years 3 to 8, may be too great for a single measure. In Study One, the mean WARP score for the low-progress readers in Year 3 was 48 WPM, with a range of 31 to 64 WPM. As the minimum score for the WARP mean for this group is above 30 WPM, it is clearly appropriate for students in Year 3 including low-progress readers. At the upper end, the data indicate that a ceiling effect became apparent after Year 5. This does not mean, however, that the WARP is not useful above Year 5. The results of Study One indicate that the WARP is appropriate for tracking older low-progress readers above Year 5. As the aim of developing a measure such as the WARP is to track the progress of low-progress readers towards functional literacy, students reading at above Year 5 level would not, in any case, be in need of such close monitoring of their reading progress.

Thus we have demonstrated that this version of the WARP comprising five parallel passages, offers a quick, reliable, and valid assessment of reading performance for low-progress readers. We have demonstrated that the number of words read correctly in the first minute correlates very highly with the measure based on the reading of the whole passage and hence is the more cost effective alternative.
Similarly, we have demonstrated excellent alternate or parallel forms reliability and good criterion validity with reading accuracy. Moreover, we have shown that the WARP is sensitive to increasing reading performance over time and to differences between average and low-progress readers.

References


**Appendix A**

**A Sample Passage from the WARP**

Last week Dad decided that buying things cost too much money. “We could make all sorts of things if I had a workshop,” he said. “Just a little one with a workbench and a few tools.”

Mum looked worried. She did not think it was a very good idea. “Great idea, darls. The only problem is that good tools cost a lot of money.” And Mum smiled. She thought that would be the end of Dad’s idea but she was wrong.

“No problem, love,” said Dad. “My friend has a shed and some tools that he wants to sell. He’ll let me have the lot for just three hundred dollars.”

Mum gulped. What could she try next?

“Why not go and get the shed and the tools right now,” she said. “Then you could fix the toaster and the dryer. And then why not look at the back gate. It keeps sticking.”

Mum went on and on. Dad looked a bit sick and then he gave in. “Perhaps not after all, darls. There’s a really good show on the television tonight. Maybe next week.”

Mum smiled. She knew that she would not hear about the shed or the tools again.

© Kevin Wheldall, 1996.

Copying prohibited without permission.
Appendix B

Directions for Administering and Scoring Passage Reading Tests (PRTs)

Administering PRTs
Give the student the copies of the appropriate PRTs that do NOT have the numbers on the right hand side of the sheet. The student sheet should not have a passage title. Do not tell the student the title of the passage.

Ask the student to read the passage:
“I’d like you to read this passage as quickly and as carefully as you can. I’ll start recording when you read the first word.”

Begin timing when the student says the first word in the passage. Mark the last word read at the end of one minute and keep timing until the student has read to the end of the passage. Note the time taken to read the whole passage.

Scoring Procedures
Use the number column to assist calculating the number of words read in one minute and for the whole passage. From these two scores subtract errors (in the manner described below) to calculate the number of words read correctly in one minute (WPM) and the number of words read correctly per minute averaged over the whole passage (WPP).

Each mispronunciation or substitution should be scored as 1 (one) error. If a word is omitted, score 1 (one) error. If several consecutive words are omitted (e.g., a line is skipped), subtract the total number of words omitted, and score 1 (one) error. If a word is inserted, do not count the inserted word, and score 1 (one) error. If the word order is reversed, score 1 (one) error.

To calculate WPM, subtract the total number of errors made in the first minute from the total number of words read in the first minute. To calculate WPP, subtract the total number of errors from the total number of words read, divide by the total time taken (in seconds), and multiply by 60.