

**Preliminary evidence for the validity of the new Test of Everyday Reading  
Comprehension**

Kevin Wheldall<sup>1</sup>, Sarah McMurtry<sup>1</sup>

<sup>1</sup> *MultiLit Research Unit, Macquarie University Special Education Centre, Macquarie University, Macquarie Park, Australia.*

The following is the post-print version of an article that has been peer reviewed and accepted for publication. It was published by Taylor and Francis on 14 November 2014. Full reference details are below.

Wheldall, K., & McMurtry, S. (2014). Preliminary evidence for the validity of the new Test of Everyday Reading Comprehension. *Australian Journal of Learning Difficulties*, 19(2), 173-178. <https://doi.org/10.1080/19404158.2014.979525>

## **Abstract**

The Test of Everyday Reading Comprehension (TERC) has recently been presented as an addition to the armoury of tests available for assessing the skills of low-progress readers.

While comparison data for students of different ages are presented together with evidence for high test reliability, there is, as yet, no published evidence for its validity. Consequently, an incidental sample of 169 low-progress readers attending a reading clinic was assessed on a battery of reading and related skills tests and also on the TERC. Insofar as the TERC was shown to correlate highly with other relevant measures, preliminary evidence for the test's validity is provided by this study.

## **Preliminary evidence for the validity of the new Test of Everyday Reading**

### **Comprehension**

While the tests available for assessing the decoding element of reading are plentiful, there are fewer tests of reading comprehension available. Moreover, due to the lack of consensus concerning the skills involved in reading comprehension, the extant range of tests of reading comprehension varies considerably in what they measure. A common method of testing reading comprehension, reading through paragraphs of text and answering questions upon completion (see, for example, the Neale analysis; Neale, 1999), can result in the test results being confounded by the influences of memory-related skills (McArthur et al., 2013). Other limitations arise from the differing formats of tests such as being allowed to see the text while answering questions as opposed to the text being covered while answering questions; the types of questions employed, including inferential, literal or both; and response format (multiple choice, open-ended questions, maze procedures, etc.) (McArthur et al., 2013). In attempting to break down and measure the constituent skills involved, testing reading comprehension has become a very difficult task.

The authors of the Test of Everyday Reading Comprehension (TERC; McArthur, Castles, Larsen, & Coltheart, 2009) argue that their test was created as an alternative to attempting to isolate particular reading comprehension skills and, instead, to test all of the different skills involved together, at the same time. It is a test of real life reading comprehension for children aged 6–13 years old using everyday reading stimuli. This approach was taken because this type of reading comprehension reflects a person's ability to cope in society. By testing everyday reading, performance on the test is also more likely to be meaningful to parents because they would have a better understanding of what their child should be able to do.

The TERC measures a student's understanding of text that he/she might encounter in everyday life. This includes text such as written invitations, text messages and cooking instructions on food packets. It was created to detect poor reading comprehension in primary school children. There are two alternate forms of the test, both with age and grade level norms. Each form comprises 10 items with 2 questions each per item, in order of difficulty.

McArthur et al. (2013) reported the reliability of the test to be adequate. In their analysis of alternate forms reliability, the authors reported a high intra-class correlation of  $r = 0.86$  with 258 students. Inter-rater reliability was also reported to be high for both forms of the test ( $r = 0.99$ ). However, no data have yet been published on the validity of the test. While it could reasonably be argued that the TERC clearly has considerable face validity, given that it is rooted in the sorts of printed materials that children (and adults) encounter and need to be able to read and understand in everyday life, there is still a need to demonstrate that the test has concurrent validity. Consequently, in this paper, we offer provisional, tentative evidence for the concurrent validity of the TERC Form A, as presented on the MOTIF website (<http://www.motif.org.au/>). Thus, the aim of this paper was to investigate how well the TERC correlates with other reading measures on an incidental sample of low-progress readers.

## **Methods**

### **Participants**

Participants comprised of an incidental sample of low-progress readers enrolled or seeking to enrol in remedial literacy programmes at a remedial reading centre in Sydney. Data from 169 students were available for use in the present study. The mean age of these participants was 9 years 10 months (118 months, ranging from 83 to 176 months). Four students were in year 1, 37 were in year 2, 30 were in year 3, 35 were in year 4, 28 were in year 5, 18 were in year 6, 11 were in year 7 and 6 were in year 8.

### **Procedure**

All participants were administered the battery of tests face-to-face in a quiet space, usually during one sitting. Trained staff employed by the centre that delivered the remedial literacy program administered the tests. The TERC was downloaded from the MOTIF website (<http://www.motif.org.au/>) set up by the test authors and was delivered according to the instructions provided with this version of the test.

Note that an earlier version of Form A of the TERC with 13 items (26 questions) was initially used in this short study. The most recent version of the test, with only 10 items and 20 questions, later became available on the MOTIF site and was subsequently used. With the permission of the test authors, the earlier administrations of the test were recoded to reflect the scoring of the published forms of the test. This was achieved by removing scores obtained on items that were removed from the more recent version of the test to recalculate the total items correct. Later, participants in this study were administered the published, 10-item version. There were two methods of calculating raw scores for the TERC. One was the total number of questions answered correctly out of 20 total questions and the other was the total number of items answered correctly out of 10 items, with 2 questions per item. Norms are provided for scores out of 20 and so these raw scores were used in the analysis.

In order to estimate the criterion-related validity of the TERC, correlations were calculated between the raw scores of the TERC and the raw scores of the other tests administered in the battery.

## **Measures**

The battery of tests administered in addition to the TERC (see 'Introduction' section) comprised the following.

Burt word reading test (Gilmore, Croft, & Reid, 1981). This test measures single word recognition via a list of words increasing in difficulty and provides equivalent reading ages

up to 13 years old. The test authors reported high test–retest reliability ( $r = 0.96–0.99$ ) and internal consistency ( $r = 0.96–0.97$ ). The test has also demonstrated adequate criterion-related validity with other tests of reading ability.

South Australian spelling test – second edition (Westwood, 2005). This test provides a spelling age estimate ranging from 6 to 15 years based on the number of correct items. The test authors reported good test–retest reliability ( $r = 0.96$ ) for most year groups. Similarly, alternate forms reliability was also demonstrated to be adequate ( $r = 0.89–0.94$ ).

Neale analysis of reading ability – third edition (Neale, 1999). This test measures reading accuracy, reading rate and reading comprehension. Students read up to six passages of text, which are timed, and errors are recorded and corrected. Following the completion of a passage, the test administrator asks a set of comprehension questions relating to the passage. The reading accuracy errors are used to calculate a reading accuracy score and correct answers to comprehension questions are used to calculate a comprehension score. A reading rate score is calculated using reading errors and the recorded time for each passage but were not included in this study. The test manual reported adequate internal consistency ( $r = 0.71–0.96$ ) and good criterion-related validity for accuracy and comprehension with the Schonell graded word reading test ( $r = 0.88–0.96$ ).

Martin and Pratt nonword reading test (Martin & Pratt, 2001). The Martin and Pratt test assesses phonological recoding ability with pseudo words presented in order of increasing difficulty. The test is normed for children aged 6–16 years old. The test has demonstrated high test–retest reliability ( $r = 0.95–0.96$ ) and good alternate forms reliability over different age groups ( $r = 0.90–0.94$ ). The test authors report good internal consistency ( $r = 0.93–0.96$ ).

Peabody picture vocabulary test – fourth edition (Dunn & Dunn, 2007). This test is a measure of receptive vocabulary and is commonly employed as a rough proxy for overall

verbal ability. The test comprises pages of four pictures each. The test administrator provides a verbal stimulus word and students are required to respond by selecting the picture that they think best represents the spoken word. The test is suitable for people aged 2–90 years old. The test manual provides evidence of high internal consistency ( $r = 0.89–0.98$ ) across age groups, grades and forms of the test, adequate alternate forms reliability (adjusted for range restriction,  $r = .0.87–0.93$ ) across age groups, and high test–retest reliability (adjusted for range restriction,  $r = 0.92–0.96$ ) across age groups. Correlation coefficients of the associations between the test and the expressive vocabulary test ( $r = 0.80–0.84$ ), the comprehensive assessment of spoken language ( $r = 0.37–0.77$ ) and the clinical evaluation of language fundamentals ( $r = 0.67–0.79$ ) provide evidence of the criterion-related validity of the test.

Wheldall assessment of reading passages (WARP) (Wheldall & Madelaine, 2013). The WARP is a curriculum-based measure of passage reading fluency for students reading at a level of years 2–5. Students read each of three initial assessment passages for 1 min and the total words read correctly in that minute (WCPM) are calculated for each passage, which are then averaged to obtain a mean score over the three passages. The test manual reported high test–retest reliability ( $r = 0.86–0.97$ ) and high internal consistency ( $r = 0.97–0.99$ ). Good criterion related validity was also reported with the Neale analysis of reading ability – revised reading accuracy ( $r = 0.78–0.87$ ) and reading comprehension ( $r = 0.67–0.72$ ) and the Burt word reading test ( $r = 0.83–0.85$ ).

## **Results and discussion**

The TERC was administered to participants as part of a battery of tests routinely administered to students entering the clinic. All data from the TERC were the result of the first administration of the test to each participant. One hundred and nine participants were also administered an additional test (the Peabody picture vocabulary test – fourth edition),

which was a test only administered to students who were not yet enrolled at the centre. Two students were not assessed on the Martin and Pratt test, and thus data for this test were available for only 167 students. See 'Measures' section for a full description of tests used.

Means and standard deviations for all measures (raw scores) are shown in Table 1, along with minimum and maximum scores. As may be seen, the obtained data sampled almost the full range of possible scores for the TERC.

The pattern of correlations showed that the TERC correlated significantly ( $p < 0.0005$ ) with all of the other measures. Correlations were high between the TERC and the Burt ( $r = 0.75$ ) (a measure of single word reading) and with both Neale reading accuracy ( $r = 0.74$ ) and Neale reading comprehension ( $r = 0.71$ ). A high correlation was also found with the WARP, a measure of reading fluency which is correlated highly with reading accuracy and reading comprehension and provides a good estimate of overall reading ability (Madelaine & Wheldall, 1999).

The TERC correlated less highly with the PPVT ( $r = 0.56$ ), a measure of receptive vocabulary sometimes employed as a proxy for listening comprehension or verbal ability more generally. There was a similar correlation between the TERC and the Martin and Pratt non word reading test ( $r = 0.55$ ). The TERC, then, appears to be measuring something similar to the other tests of reading and related skills employed and also, to a lesser degree, with listening comprehension.

Furthermore, a subset of 54 students had received a second administration of the test battery following two terms of intensive literacy instruction in the reading clinic, providing pre- and post-test data for these students. This enabled further analysis of the test as a valid measure of growth in reading comprehension. Means and standard deviations for the TERC and Neale comprehension measures for this subset of data are shown in Table 3.

First, the correlation between the pre- and post-data for the TERC measure ( $r = 0.72$ ,  $p < .0005$ ) provided further evidence for the test–retest reliability of the TERC. Although the correlation coefficient was lower than is typically acceptable as strong evidence of test–retest reliability, it should be noted that the test was administered on two occasions 6 months apart and during which intervention took place which might reasonably be assumed to have been differentially effective across participants. This supports the findings of the test authors, who reported good test–retest reliability (McArthur et al., 2013).

Results of paired t-test analyses on students for whom we had pre- and post-test data on the TERC and the other test of reading comprehension we employed (Neale analysis reading comprehension) are presented in Table 3. As the results show, students made highly significant ( $p < .0005$ ) gains on both measures of reading comprehension, TERC and Neale comprehension, both with large effect sizes (TERC  $d = 1.44$ ; Neale comprehension  $d = 1.74$ ). This provides preliminary evidence of the TERC's sensitivity to growth insofar as it appears to measure growth in reading comprehension as well as an established measure of the skill (Neale reading comprehension).

It should be noted that the incidental sample used might limit this study in terms of the generalisability of the results. The sample comprised generally low-progress readers who sought assistance at a clinic specialising in remedial literacy instruction. Thus, these findings may not necessarily apply to typically progressing readers and to socially disadvantaged students who would not have the means of accessing private tutoring. However, the range of scores obtained by the sample on the TERC indicates that the findings may not be so restricted by the literacy skills of the sample. Further research with larger, more diverse samples would be valuable for providing further evidence of the validity of the test.

The TERC shows promise as a measure of real world reading comprehension in that it correlates well with other measures of both reading and listening comprehension and appears to be a sensitive measure of growth in reading comprehension.

## References

- Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test* (4th ed.). Minneapolis, MN: NCS Pearson.
- Gilmore, A., Croft, C., & Reid, N. (1981). *Burt Word Reading Test – New Zealand revision*. Wellington: New Zealand Council for Educational Research.
- Madelaine, A., & Wheldall, K. (1999). Curriculum-based measurement of reading: A critical review. *International Journal of Disability, Development and Education*, 46, 71–85. doi:10.1080/ 103491299100731
- Martin, F., & Pratt, C. (2001). *The Martin and Pratt nonword reading test*. Melbourne: Australian Council for Educational Research.
- McArthur, G., Castles, A., Larsen, L., & Coltheart, M. (2009). *Test of Everyday Reading Comprehension (TERC)*. Retrieved from [www.motif.org.au](http://www.motif.org.au)
- McArthur, G., Jones, K., Anandakumar, T., Larsen, L., Castles, A., & Coltheart, M. (2013). A Test of Everyday Reading Comprehension (TERC). *Australian Journal of Learning Difficulties*, 18, 35–85. doi:10.1080/19404158.2013.779588
- Neale, M. (1999). *Neale analysis of reading ability* (3rd ed.). Melbourne: Australian Council for Educational Research.
- Westwood, P. (2005). *Spelling: Approaches to teaching and assessment* (2nd ed.). Camberwell: Australian Council for Educational Research Press.
- Wheldall, K., & Madelaine, A. (2013). *The Wheldall assessment of reading passages (WARP) manual*. Sydney: MultiLit Pty.

Table 1. Means, standard deviations and range (raw scores) for all measures

Measure	N	Min	Max	Mean	SD
TERC	169	1	20	13.44	4.82
Neale accuracy	169	2	88	36.04	15.59
Neale comprehension	169	0	37	11.75	6.05
Burt	169	9	92	47.56	17.60
South Australian Spelling	169	12	56	28.98	8.06
WARP	169	7	227	82.17	41.85
PPVT-IV	109	89	196	147.76	18.67
Martin and Pratt	167	3	47	21.55	9.12

Table 2. Correlations between TERC and all other measures

	TERC	Neale accuracy	Neale comprehension	Burt	South Australian Spelling	WARP	PPVT-IV
Neale accuracy	.742*						
Neale comprehension	.707*	.757*					
Burt	.752*	.909*	.720*				
South Australian Spelling	.699*	.832*	.571*	.848*			
WARP	.732*	.883*	.655*	.848*	.789*		
PPVT-IV	.560*	.558*	.605*	.585*	.458*	.528*	
Martin and Pratt	.545*	.708*	.464*	.743*	.698*	.618*	.378*

\*\*p<.0005, two-tailed

Table 3. Means and standard deviations and resultant gains (raw scores) for TERC and Neale comprehension measures.

Measure	N	Pre-test Mean (SD)	Post-test Mean (SD)	t	p	ES*
TERC	54	12.24 (5.24)	15.83 (4.17)	7.22	<.0005	1.44
Neale comprehension	54	9.76 (4.86)	14.70 (6.48)	8.54	<.0005	1.74

\*ES – effect size