

The Test of Word Reading Efficiency (TOWRE) used in an Australian context

Bruce Allen Knight and Susan A. Galletly
Central Queensland University

Abstract

Australian results on PISA 2000 (Program for International Student Assessment; OECD, 2002) show polarised reading achievement, with large proportions of strong readers (>40%) but also large proportions of weak readers (>30%). From the results it would seem that current instruction meets many readers' needs, but not those of this lower third. With reading accuracy not emphasised in Australian reading instruction in recent decades, reading accuracy difficulties may underlie this weak achievement. This paper discusses issues that underlie the optimising of reading instruction in Australia, the importance of applied research at school level in resolving these issues, and the merits of use of rigorous reading accuracy tests in this research. The Test of Word Reading Efficiency (TOWRE) is discussed as a potential test for gathering such evidence and data, and the findings from testing of 1200 students in one Queensland region are discussed.

Reading accuracy is the reading (identifying) of the spoken words corresponding to written words. It is integrally related to academic learning and progress (Adams, 1990; National Research Council, 1998). This is because, as Adams (1990, p.3) concludes, the ability to read words, quickly, accurately, and effortlessly, is critical to skilful reading comprehension.

Optimal reading accuracy instruction, as succinctly stated by Torgesen (1998), is the right instruction at the right level of intensity and duration for the right children at the right time. In Australia and internationally, there have been opposing viewpoints for three decades on what constitutes the right instruction, intensity, time and children when it comes to reading accuracy instruction. Termed the Reading Wars (Louden et al., 2006), this divisiveness can be considered to be between proponents of skills-based and meaning-based paradigms of reading development and instruction. Skills-based proponents promote skills development instruction using decontextualised words and word parts in addition to reading of meaningful texts, while meaning-based proponents insist that reading instruction must use meaningful text, and not decontextualised words and word parts (Hempenstall, 2003).

The issues which separate skills-based and meaning-based paradigms include:

- The role of English's extremely high orthographic complexity: whether it makes reading accuracy skills development essential or nigh on impossible;
- The role of reading accuracy and being able to read decontextualised words in reading development: whether reading accuracy is pivotal for reading comprehension and academic learning, or of lesser consequence;

- Reading accuracy instruction using decontextualised words and orthographic units: whether it is important, of little consequence, or even damaging to children's reading development;
- Strategies for working out unfamiliar words: whether they should primarily be reading accuracy, or language comprehension (semantic and syntactic);
- The rigour of experimental research findings: whether they are valid and reliable, or not; and,
- Reading accuracy assessment using decontextualised words: whether such assessment is highly important, or unnecessary and irrelevant.

Australian instructional supports for reading instruction have built from the meaning-based paradigm since the 1970s. Reading accuracy instruction has not been emphasised in Australian reading instruction and curriculum documents (Hempenstall, 2003), and reading accuracy has not been assessed routinely in Australian schools (de Lemos, 2001).

In recent years, there have been indicators that reading instruction may be inadequate for lower-achieving readers (OECD, 2002, 2004; Education Queensland, 2000). Australian schools and teachers seem to consider current reading instruction inadequate in major ways; for example, in Education Queensland's (2000) Literate Futures review, the most common concerns of schools and teachers were inadequacies of current reading instruction. The PISA 2000 findings additionally suggest differential adequacy of current instruction. Whereas many countries showed consistently high achievement or consistently low achievement, Australia showed both, with 42% of students in the top two categories, and 31% at or below the lowest two (OECD, 2002). In addition to its very wide interquartile range, Australia

had the widest reading ability gap of all nations between students who read independently beyond school and those who don't. Australian PISA 2003 results show little difference to those of PISA 2000 (OECD, 2004). With little emphasis on reading accuracy in recent decades, inadequate reading accuracy may underlie this polarised achievement. Because of its pivotal role in effective reading comprehension, successful reading accuracy achievement may be a gateway separating students with efficient reading accuracy, who go on to flourish in literacy, from those with inadequate reading accuracy, who continue to struggle across the school years.

It is highly likely that there is a common basis to polarised reading achievement and academic divisiveness on how reading should be taught, namely the difficulties created by English's complex orthography (26 letters making 44 common sounds using hundreds of spelling patterns) for many children learning to read in English-speaking nations, and the enormous challenges of developing instruction effective for all readers in these nations. In European nations, which maintain simple orthographies (spelling systems with few rules beyond each letter having one sound), almost all students master reading accuracy and within their first school year (Galletly & Knight, 2004; Seymour, Aro, & Erskine, 2003). Reading accuracy instruction there is not just brief, it is also uncomplicated. English-text reading accuracy instruction, in contrast, is complex and lengthy, with even healthy progress readers taking many years to master reading accuracy (Hanley, Masterson, Spencer, & Evans, 2004).

In developing the right instruction for the right students (Torgesen, 1998), it is likely the way forward lies in gathering data and conducting research to establish the patterns of reading development of healthy progress and at-risk readers, and principles of instruction which support optimally differentiated instruction to meet the different instructional needs of different groups of students.

The report of Australia's National Inquiry into the Teaching of Literacy emphasises that "teachers must have access to, understand, know, and be able to use teaching strategies that have consistently been shown from evidence-based research findings to be demonstrably effective" (DEST, 2005, p. 38) and that these strategies must be effective with both healthy progress and at-risk students. There are thus needs for evidence-based research to establish these strategies. Much of this research should be in school and classroom contexts, for several reasons. Firstly, research is consistently pointing to classroom learning effects as a major factor impacting student learning and development (Cuttance, 2001; Loudon et al., 2006). Secondly, little experimental

reading research has been replicated in applied (school and classroom) settings (Stanovich, 2004; Swanson, Hoskyn, & Lee, 1999) and, because of the impact of classroom learning effects on effectiveness of instruction, this must be done before definitive conclusions regarding classroom instruction can be made. Thirdly, current experimental research is focused very much on at-risk and delayed readers, and there is insufficient research on healthy progress readers to support decision-making regarding these students' reading accuracy instructional needs (Pressley, 2003). Fourthly, research findings are often focussed on group response to intervention, rather than on the results of different subgroups of learners (Pressley, 2003). In PISA 2000, Australia ranked as sixth highest of the 32 nations, and significantly below only Finland. It is only when the results of different subgroups are considered that the information available in the PISA data is made clear. Australian reading achievement is not consistent across high and low achievers. Australia has 30 per cent of weak readers whose needs aren't being met, and large proportions of good readers who are reading extremely well. Finally, relatively little research has focused on important specifics of English reading accuracy development and instruction, such as:

- Patterns of reading development across the primary school years, showing development of skill with different word types, e.g., high frequency words, regular single syllable and multi-syllabic words using English's 30 common vowel graphs, and words with different spelling patterns;
- Patterns of generalisation of learning, e.g., how healthy progress and at-risk readers transfer reading accuracy skills learned in decontextualised situations into skills used in authentic reading;
- Patterns of reading development with different instructional supports, e.g., the impact on healthy progress and at-risk readers of reducing adult:child ratios for reading instruction to 12:1, as strongly recommended by school associations (Learning First Alliance, 2000).

Reflective assessment, planning and instruction are a balancing act, an integrated triad supporting the development of optimal instruction. All three are necessary. Optimal assessment will not only provide valuable information to inform planning and instruction, but also use minimal instruction time in doing so. Used with integrity, reading accuracy tests are instruction-neutral. They provide data on students' stages of reading development, and indicate strengths and weaknesses, but do not dictate the precise form instruction should take, e.g. decontextualised or embedded instruction, or more time spent reading. Being instruction-neutral, the tests support monitoring of the differential effectiveness

of particular forms of instruction with different groups of students, through monitoring student progress over the school terms in which the instruction is used.

Such assessment supports teachers' fine-tuning of classroom instruction to achieve optimal progress. It also supports research focused on developing principles of optimal differentiated instruction. As an exemplar, it seems likely that the most common reading instruction strategy used with Australian readers is time spent reading books of appropriate challenge, that is, books with text at instructional and independent reading level. It is also likely that this strategy is extremely effective (OECD, 2002, 2004). It would seem highly useful to conduct research to establish the differential effectiveness, for students with different progress rates, and at different stages of reading development, of:

- Time spent reading books of appropriate challenge;
- Systematic decontextualised reading accuracy instruction;
- Specific combinations of these strategies.

Use of rigorous reading accuracy tests in school-level research has potential to answer such issues.

In recent years, USA researchers have developed a new genre of reading accuracy tests, characterised by rigour and rapid-use. The *Test of Word Reading Efficiency* (TOWRE; Torgesen, Wagner and Rashotte, 1999) is one such test. TOWRE takes less than five minutes to administer and process results, and can be administered by teachers and aides. It measures efficiency of word reading, that is, the interaction of speed and accuracy aspects of reading accuracy. TOWRE has two subtests, and each has two alternate forms:

1. Sight Word Efficiency (SWE) measures efficiency of reading real words, using a single list of 104 words of one to four syllables; and,
2. Phonemic Decoding Efficiency (PDE) measures efficiency of reading pseudowords (non-words), using a single list of 63 pseudowords of one to three syllables.

The words used include regular and irregular words, and use common and less common vowels, vowel

digraphs, and orthographic units. The same one-page stimulus list for each subtest is used for all readers, regardless of age and skills. The test has USA norms for readers aged 6-0 to 24-11, and established high levels of reliability (≥ 0.94) and validity (0.89-0.85-0.94), such that users can be confident with test results.

SWE results indicate students' familiarity and efficiency with real words, while PDE results show efficiency with unfamiliar words. Ability to read and write pseudowords is strongly linked to ability to read unfamiliar words in independent reading, and to spelling progress (Torgesen et al., 1999), thus PDE provides an additional perspective on students' enthusiasm for independent reading, and spelling skills. Limited diagnostic decision-making can also be conducted from analysis of student achievement and errors on different graphemes and word types.

The TOWRE is a screening test which can quickly identify healthy progress students who are responding well to current instruction, and those who will benefit by closer monitoring of achievement and response to changes in instruction. In school usage, TOWRE provides data for monitoring the achievement and progress of individual students. This data can also be used in school and system monitoring of achievement and instructional effectiveness, and in reading research focused on optimising instruction.

The current study explored the use of TOWRE in an Australian context. This paper focuses on exploring the usefulness of TOWRE in gathering data on reading accuracy achievement, and the types of knowledge that can be gained from TOWRE data. TOWRE data options include raw scores, reading ages, grade equivalents, standard scores and percentiles. In the current study, an additional researcher-developed option, Age Gap (Chronological Age – Reading Age) was trialled, because of the flexibility this data option offers for gathering comparable data despite testing at different times of year. The current study explored the use of TOWRE in an Australian context. It is part of a larger reading accuracy study focused on measuring reading accuracy

Table 1: Sample characteristics including number of schools and subjects in each year level (Y).

Years 1-4				Years 5-8			
Year	Schools	Classes	Subjects	Year	Schools	Classes	Subjects
Y1	4	8	133	Y5	2	4	179
Y2	5	8	169	Y6	2	4	146
Y3	5	9	187	Y7	2	4	105
Y4	1	2	49	Y8	1	6	237

achievement using different reading accuracy tests. this paper focuses on exploring the usefulness of TOWRE in gathering data on reading accuracy achievement, and the types of knowledge that can be gained from TOWRE data.

Method

In this study, volunteer schools and teachers in a central Queensland district were invited to work with the researchers to assess their students using TOWRE. Testing was conducted by teachers, teacher aides, and volunteer education undergraduates trained by and working with the researchers. Five primary schools and one high school participated, with the sample comprising 1205 students in Years 1-8 (see Table 1). Gender ratios were similar with non-significant differences, except in Year 1 where 60% were boys. All students speak English as their first language.

The likelihood of results being impacted by school and classroom effects decreases when more schools and classes are used, thus Year 1-3 data, from four and five schools, is more likely representative of regional achievement than Year 4-8 data (see Table 1). With just one school for Year 4, Year 4 data was excluded from data processing, due to strong likelihood of school and classroom effects. Year 8 data was included, despite being from just one school, as school effects were minimised by Year 8 being the first year of high school, and this school drawing their population from numerous primary schools.

This study is preliminary exploratory research, looking for trends which can be explored more rigorously in future research, rather than research for definitive findings.

Results and discussion

Comparability of data options

Correlation coefficients were used to compare the power of five data options: Raw Score, Reading Age, Age Gap, Grade Equivalent, and Standard Scores (TOWRE percentiles derive from standard scores, so are not included here). All correlations were highly significant ($p < 0.001$), for each subtest (SWE and PDE) and at each year level, for whole year levels, and for separate male and female cohorts. Coefficients were all above 0.7, with the vast majority above 0.9. Given such strong correlations, it seems appropriate for schools and systems to use whichever of these data options they consider most useful for their consideration of student performance.

Mean achievement

Mean reading accuracy achievement for all year levels was in the USA-norms average range with all mean standard scores above 95. Mean reading ages for both SWE (real words) and PDE (non words) were strongly correlated with students' chronological ages ($p < 0.001$), with evidence of healthier achievement in Years 1-3, and a levelling off of achievement from Years 5 and 6 (see Figure 1).

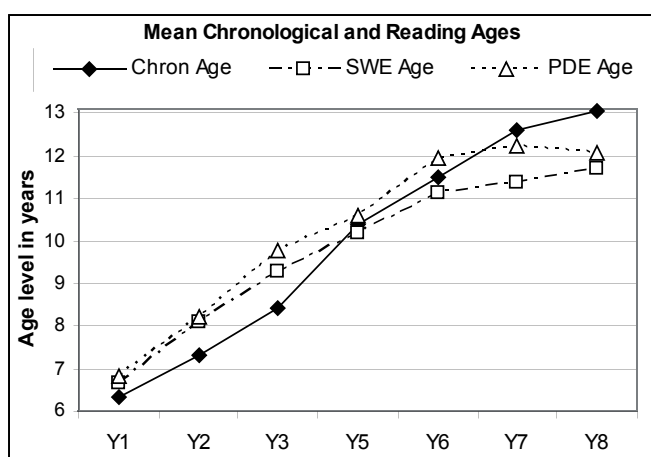
Gender

Female means were higher than male means for SWE and PDE at all year levels, except PDE in Years 3 and 5, where means were equal. While not all these differences were significant (Galletly & Knight, submitted), a trend worth further investigation is evident.

Spread of achievement

In the current study, deciles were used to consider achievement by different groups of students (see Figure 2).

Figure 1: Mean SWE reading age, PDE reading age and chronological age across year levels.

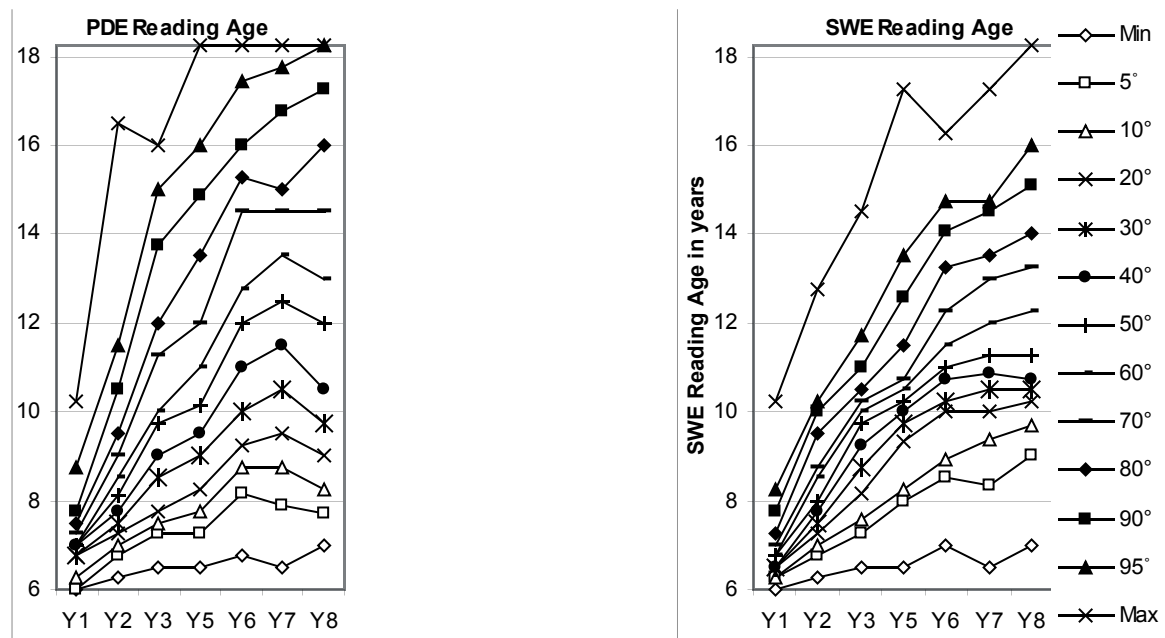


Mean Chronological and Reading Ages								
		Y1	Y2	Y3	Y5	Y6	Y7	Y8
SWE Age	x	6.6	8.1	9.3	10.2	11.1	11.4	11.7
	sd	0.6	1.2	1.4	1.6	1.9	2.0	2.2
PDE Age	x	6.8	8.2	9.8	10.6	12.0	12.2	12.1
	sd	0.7	1.5	2.4	2.8	2.9	3.0	3.4
Student Age	x	6.3	7.3	8.4	10.4	11.5	12.6	12.8
	sd	0.3	0.4	0.4	0.4	0.4	0.4	0.4

Deciles create groups of one-tenth: in a class of 30 students, there would be three students in each group. It is possible deciles will prove a useful statistical grouping for teachers and schools. Where finer-grained information is needed, smaller groupings can be considered, e.g., because the most marked spread was in the top and bottom tenths, the 5th and 95th percentiles were also included. These show that extreme results are in the highest and lowest 5%, not 10%, of students.

Flattening of achievement from Year 6 is evident in the lower 50% (SWE) and 70% (PDE) of readers. Also visible in the data is the very wide range of ability present in the classes tested, and likely to be present beyond the current sample. In Years 1-3, the range increases from four to eight years, while from Year 5, 10 years of reading ability are present. Statistically, this wide range simply reflects acceptable variability within a normal distribution. In school and classroom contexts, however,

Figure 2: Spread of reading age achievement across year levels using decile groups.



	SWE Reading Age							PDE Reading Age						
	Y1	Y2	Y3	Y5	Y6	Y7	Y8	Y1	Y2	Y3	Y5	Y6	Y7	Y8
Max	10	12.5	14.3	17	16	17	18	10	16.3	15.8	18	18	18	18
95°	8	10	11.5	13.3	14.5	14.5	15.8	8.5	11.3	14.8	15.8	17.2	17.5	18
90°	7.5	9.8	10.8	12.3	13.8	14.3	14.8	7.5	10.3	13.5	14.6	15.8	16.5	17
80°	7	9.3	10.3	11.3	13	13.3	13.8	7.3	9.3	11.8	13.3	15.1	14.8	15.8
70°	6.8	8.5	10	10.5	12	12.8	13	7	8.8	11	11.8	14.3	14.3	14.3
60°	6.5	8.3	9.8	10.3	11.3	11.8	12	6.8	8.3	9.8	10.8	12.5	13.3	12.8
50°	6.5	7.8	9.5	10	10.8	11	11	6.8	7.9	9.5	9.9	11.8	12.3	11.8
40°	6.3	7.5	9	9.8	10.5	10.6	10.5	6.8	7.5	8.8	9.3	10.8	11.3	10.3
30°	6.3	7.3	8.5	9.5	10	10.3	10.3	6.5	7.3	8.3	8.8	9.8	10.3	9.5
20°	6.3	7	7.9	9.1	9.8	9.8	10	6.5	7	7.5	8	9	9.3	8.8
10°	6	6.8	7.3	8	8.7	9.2	9.4	6	6.8	7.3	7.5	8.5	8.5	8
5°	6	6.5	7	7.7	8.3	8.1	8.8	5.8	6.5	7	7	7.9	7.6	7.5
Min	5.8	6	6.3	6.3	6.8	6.3	6.8	5.8	6	6.3	6.3	6.5	6.3	6.8

it represents major instructional challenge, with many lower-achieving students having needs for instructional intensity that may well be beyond that available with current school funding. These needs may include:

- Use of several separate ability-groups of weak readers in each year level, given the wide range of ability in the lower 30% of achievers;
- Instructional time additional to usual classroom literacy sessions, for each group, focused specifically on reading accuracy development;
- Instruction sessions scheduled three times per week, and preferably daily for students below the 10th percentile; and,
- Adult:child ratios for these sessions reduced from the usual 1:25 of upper primary school to 1:4 for most students and 1:1 for students not making effective progress in small group instruction.

Relationship with academic achievement

Correlations of PDE and SWE standard scores with students’ rasch scores on National Benchmark tests (Queensland’s Year 3, 5, 7 Tests) of Reading, Writing, and Spelling were significant for each TOWRE subtest (SWE and PDE) in each year level. Correlation coefficients ranged from 0.3 to 0.8. Significance levels for real word reading (SWE) were all highly significant ($p<0.001$), while some nonword reading (PDE) correlations were slightly less significant ($p<0.01$).

An interesting trend emerged across year levels with regard to strength of relationship of SWE standard scores with students’ national Year 3 test results, with a steady increase in correlation coefficients across the year levels (see Table 2).

While it has long been acknowledged that early reading accuracy achievement predicts academic achievement (Torgesen, 1998), it may equally be that early academic achievement predicts reading accuracy achievement.

Conclusion

This paper has discussed issues that underlie the optimising of reading instruction in Australia, the importance of applied research at school level in resolving these issues, and the merits of use of rigorous reading accuracy tests in this research. It has described the characteristics of TOWRE, a rigorous rapid-use reading accuracy tests which can be used across the school years. It has explored the results of an Australian study which used the TOWRE to assess students in consecutive year levels, and provided examples of the different types of information which can be derived from use of TOWRE.

Tests such as TOWRE would seem to have strong potential for Australian use. School-level use of TOWRE will support effective differentiation of classroom reading instruction. Additionally, TOWRE data has strong potential for use in applied school-level research focused on developing rigorous principles of effectively differentiated reading instruction. At the current time, there is much knowledge to be built in Australia to establish teaching strategies based on findings from rigorous, evidence-based research, which are effective not just with at-risk and delayed students, but with all students (DEST, 2005, p. 39). Rigorous rapid-use reading accuracy tests such as TOWRE would seem to have a useful role in this knowledge-building.

Acknowledgement

The research project discussed in this paper was made possible through the support of Education Queensland (Mackay-Whitsunday region), Catholic Education Office (Rockhampton diocese), and teacher researchers in Mackay region.

Table 2: Correlation coefficients for SWE standard score for each year level, showing increased strength of relationship in older students (all were highly significant, at $p<0.001$).

Year level in 2005	Year 3 National Test		
	Reading	Spelling	Writing
Year 3	0.422	0.508	0.400
Year 5	0.654	0.661	0.525
Year 6	0.726	0.736	0.660
Year 7	0.842	0.822	0.726

References

- Cuttance, P. (2001). The impact of teaching on student learning. In K. J. Kennedy (Ed.), *Australian College of Education Yearbook 2001: Beyond the rhetoric: Building a teaching profession to support quality teaching*. Canberra: Australian College of Education.
- de Lemos, M. M. (2001). *Reading standards up or down: What do the test norms say?* Paper presented at 2001 AARE Conference.
- DEST. (2005). *Teaching reading: Report and recommendations: National inquiry into the teaching of literacy*. Canberra: Department of Education, Science and Training.
- Education Queensland. (2000). *Literate futures: Report of the literacy review for Queensland state schools*. Brisbane: State of Queensland.
- Galletly, S. A., & Knight, B. A. (2004). The high cost of orthographic disadvantage. *Australian Journal of Learning Disabilities*, 9(4), 4-11.
- Hempenstall, K. (2003). The three-cueing system: Trojan horse? *Australian Journal of Learning Disabilities*, 8(2), 15-23.
- Hanley, J. R., Masterson, J., Spencer, L. H., & Evans, D. (2004). How long do the advantages of learning to read a transparent orthography last? An investigation of the reading skills and incidence of dyslexia in Welsh children at 10 years of age. *The Quarterly Journal of Experimental Psychology*, 57.
- Learning First Alliance. (2000). *Every child reading: An action plan*. <http://www.learningfirst.org/publications/reading/> (Retrieved 26 Sept, 2006).
- Louden, W., Rohl, M., Barratt-Pugh, C., Brown, C., Cairney, T., Elderfield, J., et al. (2006). *In teachers' hands: Effective literacy teaching practices in the early years of schooling*. Canberra: DEST.
- National Research Council. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- OECD. (2002). *PISA in brief from Australia's perspective*. Canberra: ACER; Organisation for Economic Co-operation and Development (OECD) Program for International Student Assessment (PISA).
- OECD. (2004). *PISA 2003 in brief from Australia's perspective*. Canberra: ACER; Organisation for Economic Co-operation and Development (OECD) Program for International Student Assessment (PISA).
- Pressley, M. (2003). A few things reading educators should know about instructional experiments. *Reading Online*, www.readingonline.org/articles (Retrieved 24 Oct, 2006).
- Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94, 143-174.
- Stanovich, K. E. (2004). *How to think straight about psychology* (7th ed.). Boston: Allyn & Bacon, Pearson Education, Inc.
- Swanson, H. L., Hoskyn, M., & Lee, C. (1999). *Interventions for students with learning disabilities: A meta-analysis of treatment outcomes*. New York: The Guilford Press.
- Torgesen, J. K. (1998). Catch them before they fall: Identification and assessment to prevent reading failure in young children. *LD Online, Spring/Summer*, www.ldonline.org/ld_indepth/reading/torgeson_catchthem.html (Retrieved 6 Feb, 2006).
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency (TOWRE)*. Austin, Texas: Pro-Ed, Inc.