Nonword Reading Tests:
A Review of the Available Resources

Danielle Colenbrander[1,2], Lyndsey Nickels[2] and Saskia Kohnen[1,2]

[1] ARC Centre of Excellence in Cognition and its Disorders (CCD), Macquarie University, Sydney, Australia
[2] MULTILIT Research Unit, Sydney, Australia

Response to Intervention (RTI) models of reading instruction have received much attention in the literature (Fuchs, Mock, Morgan, & Young, 2003; Hosp & Ardoin, 2008; Justice, 2006). Such models call for frequent and high-quality assessment of students' skills. One skill that is vital in the process of learning to read is the ability to sound out letters and blend these sounds together to produce a word. According to Dual Route models of reading (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), these skills comprise the sublexical route to reading. This is best assessed by tests of nonword reading. In this article, 17 tests of nonword reading were reviewed in light of RTI and Dual Route models. The aim of the review was to determine the best available nonword reading tests for use at the Tier One and Tier Two levels of intervention, and the best available nonword reading test for diagnostic (Tier Three) purposes. The review determined that several good-quality tests of nonword reading, suitable for assessing the general functioning of the sublexical route to reading (at the Tier One and Tier Two level), are available, but that no test of nonword reading is available that fulfils all of the desirable criteria for a Tier Three, diagnostic assessment.

Keywords: reading measures, phonics, remedial reading

Much has been written recently about the need for high-quality assessments of early reading (see, e.g. Hempenstall, 2009; Rathvon, 2004; Westwood, 2009). There has also been discussion on the different types of reading assessment and their purpose (Catts, Petscher, Schatschneider, Bridges, & Mendoza, 2009; Coyne & Harn, 2006; Knutson, Simmons, Good, & McDonagh, 2004). Much of this discussion has focused on assessment as it relates to Response to Intervention (RTI) models of reading intervention.

RTI models propose (at least) three tiers of reading instruction, each more intensive than the last (Catts et al., 2009; Justice, 2006). The first tier involves the delivery of evidence-based best-practice reading instruction to all children in the classroom (Justice, 2006). At this level, regular assessment of reading skills is required to ensure that children are making sufficient progress (Catts et al., 2009; Justice, 2006; Knutson et al., 2004). Because Tier One assessments must be administered to a large number of children, it is important that they can be completed in a short amount of time

(Hosp & Ardoin, 2008). For this reason, Tier One assessments will necessarily be less detailed than other types of assessment (Hosp & Ardoin, 2008).

Tier Two instruction involves more intensive delivery of evidence-based reading instruction to children who are not progressing sufficiently under Tier One instruction (Justice, 2006). Once again, regular, relatively quickly administered assessments are required to identify whether students are progressing sufficiently (Justice, 2006; Knutson et al., 2004). Students who do not progress sufficiently at this level move on to the third tier of instruction.

The third tier involves more tailored, individualised and ongoing support (Justice, 2006).[1] In order to tailor intervention to a child's individual needs, assessments such as those used in Tiers One and Two may no longer be sufficient (Hosp & Ardoin, 2008; Knutson et al., 2004). At this level, more in-depth assessments of particular reading skills are required (Hosp & Ardoin, 2008; Knutson et al., 2004). This type of assessment is often referred to as 'diagnostic assessment'.

In school settings, the vast majority of assessment will take place at the Tier One and Tier Two level. In clinical settings, diagnostic assessments are likely to be most frequently required. In research settings, Tier One and Tier Two-type assessments may be required to identify suitable participants and to compare research participants to their peers, while diagnostic assessments may be required to gain specific information required for designing interventions and treatment programs. Thus, both types of assessment have important roles in the reading intervention process (Hosp & Ardoin, 2008; Justice, 2006).

In order to know how best to help a child who is struggling to learn to read, it is necessary to determine which component reading skills are sources of difficulty for the child (Hempenstall, 2009; Kame'enui et al., 2006). This requires assessment of the child's progress in each component skill relative to their peers (Tier One and Tier Two-type assessment). Having identified which component skill (or skills) is the source of the child's difficulties, however, further diagnostic assessment of these skills is required in order that an instructional program can be tailored precisely to the child's needs.

In this article, we outline criteria necessary for determining the quality of a nonword reading test in light of Dual Route and RTI models of reading assessment. We then review available tests of nonword reading on the basis of these criteria.

Reasons for Reading Failure

A child could fail to read at the level of his or her peers due to problems with any component of the reading system; for example, they might have problems recognising letters or in storing the representations of words (Jackson & Coltheart, 2001). However, a child might also have problems with skills external to the reading system, such as vocabulary or listening comprehension skills (Department of Education, Science and Training [DEST], 2005; Rathvon, 2004). Hence, it is important to assess all of these component skills to identify the source of a child's reading problems (for further discussion, see Hempenstall, 2009; DEST, 2005). Nevertheless, research has shown that most individuals with developmental reading disabilities present with a phonologically based deficit (Rack, Snowling, & Olson, 1992) that tends to manifest itself as a problem acquiring grapheme–phoneme correspondences (Jackson & Coltheart, 2001). It is the assessment of this skill that will be the focus of our discussion.

The Dual Route Model of Reading and Acquisition of Grapheme–Phoneme Correspondences

There is empirical evidence for the existence of two routes for reading, even in young readers (Castles & Coltheart, 1993). Dual Route models of reading propose that, at the word level, reading takes place by means of the combination of these two routes: the lexical route, which incorporates a 'mental dictionary' of sight words; and the sublexical route,[2] which is thought to contain grapheme–phoneme correspondences (letter–sound rules) that can be used to sound out regularly spelled words and novel letter strings (Jackson & Coltheart, 2001; see Figure 1).

The successful functioning of this sublexical route is usually assessed by testing a child's ability to sound out nonwords (for example *pofe* and *bep,* sometimes called pseudowords). Nonwords provide the best test of this sublexical route as they cannot be read using the 'mental dictionary' of sight words and therefore successful reading of the nonword will always reflect the use of grapheme–phoneme correspondences (Jackson & Coltheart, 2001).

Determining that a child can read fewer nonwords accurately than his/her peers will identify the presence of a problem that may be alleviated by more intensive instruction at the Tier One or Two level. However, diagnostically, in order to design the most appropriately targeted intervention, it is beneficial to determine which *specific* grapheme–phoneme correspondences cause the child difficulty: if a child has already mastered some grapheme–phoneme correspondences, it does not make sense to teach the child these correspondences, but rather intervention should target only the correspondences the child has not yet mastered.

We therefore review existing tests of nonword reading in order to determine:

- the best available nonword reading tests for use at the Tier One and Two level — in other words, best for identifying the existence of a problem acquiring grapheme–phoneme correspondences
- the best available nonword reading tests for diagnostic purposes — in other words, best for identifying which grapheme–phoneme correspondences a child knows.
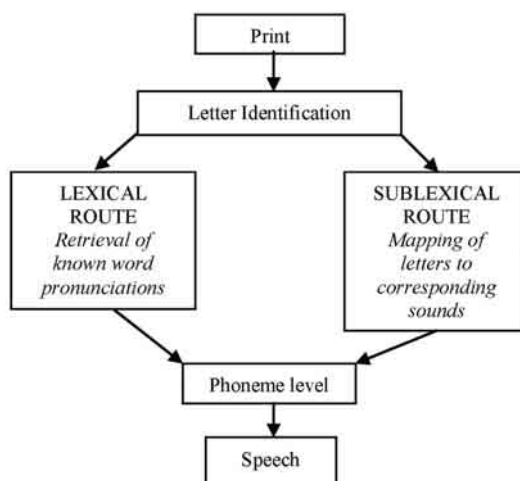

FIGURE 1
Dual Route model.

Danielle Colenbrander, Lyndsey Nickels and Saskia Kohnen

Review of Nonword Reading Tests

Because the focus of this review is on developmental reading disorders, only tests of nonword reading that were designed for use with primary school children were reviewed. These include standardised tests, placement tests from specific programs of instruction and criterion-referenced tests freely available on the internet. In the case of tests that contain subtests assessing nonword reading, only this subtest was reviewed and not the test as a whole.

Several databases were searched in order to locate tests of nonword reading. These included the EBSCOhost® Tests in Print Database and the Ovid® Mental Measurement Yearbook. Various websites were also searched, including the website of the Buros Institute of Mental Measurements (http://www.unl.edu/buros/) and the Australian Council for Educational Research (http://www.acer.edu.au/). A Google® search was also undertaken. The search terms used were 'nonword', 'pseudoword' and 'test of early reading'. Additionally, lists of reading tests, such as the alphabetical listing on the Buros website (http://www.unl.edu/buros/) were manually searched. Names of tests were also obtained from books and articles on reading assessment (e.g., Hempenstall, 2009; Rathvon, 2004; Westwood, 2009).

The search yielded 23 tests. Of these tests, 17 are reviewed (see Table 1). The authors could not obtain copies of three tests (see Appendix A). Three further tests contained a nonword subtest derived from another test that had already been reviewed, and were therefore not reviewed separately (see Appendix A).

TABLE 1

Nonword Reading Tests Reviewed in this Article

Test Name	Authors
Castles & Coltheart Reading Test 2	Castles, Coltheart, Larsen, Jones, Saunders, & McArthur, 2009
Consortium on Reading Excellence Phonics Survey, 2nd edition	Consortium on Reading Excellence (CORE), 2008
Dynamic Indicators of Basic Literacy Skills (DIBELS), 6th edition	Good & Kaminski, 2010
The Dyslexia Screening Test-Junior	Fawcett & Nicolson, 2004
Graded Nonword Reading Test	Snowling, Stothard, & McLean, 1996
Illinois Test of Psycholinguistic Abilities, 3rd edition	Hammill, Mather, & Roberts, 2001
MULTILIT Word Attack Skills Placement Test	MULTILIT, 2007
Nonword Decoding Test	Turner, 2003
Nonword Reading Test	Martin & Pratt, 2001
Nonword Reading Test	Wren, 2003
Phonological Assessment Battery	Frederickson, Frith, & Reason, 1997
Queensland University Inventory of Literacy	Dodd, Holm, Oerlemans, & McCormick, 1996
Sutherland Phonological Awareness Test-Revised (SPAT-R)	Neilson, 2003
Test of Word Reading Efficiency	Torgesen, Wagner, & Rashotte, 1999
Wechsler Individual Achievement Test, 2nd Australian edition	(The Psychological Corporation, 2002)
Woodcock Johnson III	Woodcock, McGrew, & Mather, 2001
Woodcock Reading Mastery Tests -Revised	Woodcock, 1998

Criteria for Determining the Quality of All Nonword Reading Tests

There are several criteria that need to be met for a nonword reading test to be judged as adequate. First, as with any assessment, it is important that a nonword test has good reliability and validity (Anastasi & Urbina, 1997). As this aspect is covered in detail elsewhere (Buros Institute of Mental Measurements, 2009; Rathvon, 2004), it will only be commented on briefly here, and indeed the majority of tests reviewed have sufficient levels of reliability and validity (see Table 2).

In addition to concerns of reliability and validity, a nonword reading test needs to contain a sufficient number of items of varying levels of difficulty so that it is sensitive enough to detect differences in reading skills among the youngest and least skilled readers assessed, as well as among the oldest and most skilled readers. In other words, a good nonword reading test should not be subject to floor and ceiling effects (Hempenstall, 2009; Rathvon, 2004). In practice these are extremely difficult to avoid. It is particularly challenging to come up with items that are sensitive to differences in skill among those with very limited reading ability (Rathvon, 2004). Conversely, it is difficult to come up with nonwords that are sensitive to differences in skill among those with good reading ability, because there will come a point where the majority of skilled readers can successfully read most nonwords. Only a few of the tests reviewed were not subject to floor or ceiling effects — The Martin and Pratt Nonword Reading Test (Martin & Pratt, 2001), The Dyslexia Screening Test (Fawcett & Nicolson, 2004), and the Queensland Inventory of Literacy (Dodd, Holm, Oerlemans, & McCormick, 1996); see Table 2.

Practical considerations are also important. For example, a good nonword reading test should be as quick and easy to administer as its purpose allows; there is a greater necessity for Tier One and Tier Two tests to be as short as possible than there is for diagnostic tests (Hempenstall, 2009). The majority of the tests reviewed were relatively quick to administer, with administration times of 10 minutes or less and scoring times of approximately 5 to 10 minutes (see Table 2 and Appendix C).

A stopping rule, where the test is discontinued after a certain number of errors, may decrease administration times (Castles et al., 2009) and may also reduce frustration for children who have difficulty with the task (Rathvon, 2004). However, a stopping rule will reduce the amount of information obtained about a child's reading performance, as not all items will be sampled. Thus, a stopping rule is less appropriate for a diagnostic test where systematic assessment of each and every grapheme–phoneme correspondence is required. The majority of the tests reviewed contained a stopping rule (see Table 2).

Tier One and Tier Two assessments require that nonword reading ability of an individual child is compared to that of other children of the same age or, even better, to children with the same number of years of school instruction. Hence, tests also need to contain recent and representative norms (Anastasi & Urbina, 1997). Clearly, for a test used in Australia it is best to have Australian norms. This does not mean that tests normed overseas cannot be used; however, comparisons to overseas norms need to be made with caution. Five of the tests reviewed contained Australian norms — the Martin & Pratt (Martin & Pratt, 2001), the Queensland University Inventory of Literacy (QUIL; Dodd, Holm, Oerlemans, & McCormick, 1996), the CC2 (Castles et al., 2009), the WIAT-II Australian (The Psychological Corporation, 2002) and the SPAT-R (Neilson, 2003), see Table 2.

Another practical concern is that the pronunciation of nonwords can sometimes be ambiguous, affecting a test's interrater reliability (Rathvon, 2004). It is therefore impor-

TABLE 2

Summary of Nonword Reading Test Review: General Characteristics

Test	Information on reliability/validity	Floor/ceiling effects	Easy to administer	Stopping rule	Recent norms (≤ 10 years)	Australian norms	Pronunciation guide	Link to theory
Castles & Coltheart Reading Test 2 (Castles et al., 2009)	No	Yes (floor)	Somewhat	Yes	Yes	Yes	No	Yes
Consortium on Reading Excellence Phonics Survey Consortium on Reading Excellence, 2008)	No	Ceiling effect criterion referenced	Yes	Yes	Criterion referenced	No	No	No specific theory
Dynamic Indicators of Basic Literacy Skills (DIBELS), 6th edition (Good & Kaminski, 2010)	Available 2011	Ceiling effect criterion referenced	Somewhat	Yes	Criterion referenced	No	No	Yes
The Dyslexia Screening Test–Junior (Fawcett & Nicolson, 2004)	Yes	No	Somewhat	Yes	Yes	No	No	Evidence controversial
Graded Nonword Reading Test (Snowling, Stothard, & McLean, 1996)	Yes	Yes (floor)	Yes	Yes	No	No	Yes	Yes
Illinois Test of Psycholinguistic Abilities, 3rd edition (Hammill, Mather, & Roberts, 2001)	Yes	Yes (floor/ceiling)	Somewhat	Yes	No	No	No	No recent evidence
MULTILIT Word Attack Skills Placement Test (MULTILIT, 2007)	No	Ceiling effect criterion referenced	Yes	Yes	No	No	No	Yes
Nonword Decoding Test (Turner, 2003)	Yes	Cannot determine	Yes	Yes	No	No	Some items	No specific theory
Nonword Reading Test (Martin & Pratt, 2001)	Yes	No	Yes	Yes	No	Yes	Yes†	No specific theory
Nonword Reading Test (Wren, 2003)	No	Cannot determine	Yes	No	No	No	No	Yes
Phonological Assessment Battery (Frederickson, Frith, & Reason, 1997)	Yes	Yes (floor/ceiling)	Yes	Yes	No	Yes	Yes	Yes
Queensland University Inventory of Literacy (Dodd, Holm, Oerlemans, & McCormick, 1996)	Yes	No	Somewhat	No	No	Yes	Some items	No specific theory

(continued)

TABLE 2 (CONTINUED)

Summary of Nonword Reading Test Review: General Characteristics

Test	Information on reliability/validity	Floor/ceiling effects	Easy to administer	Stopping rule	Recent norms (≤ 10 years)	Australian norms	Pronunciation guide	Link to theory
Sutherland Phonological Awareness Test–Revised (Neilson, 2003)	Yes	Yes (floor)	Somewhat	Yes	Yes	Yes	Some items	No specific theory
Test of Word Reading Efficiency (Torgesen, Wagner, & Rashotte, 1999)	Yes	Yes (floor/ceiling)	Somewhat	No	No	No	Yes	Yes
Wechsler Individual Achievement Test–2nd Australian edition (The Psychological Corporation, 2002)	Yes	Yes (floor/ceiling)	Yes	Yes	Yes	Yes	Yes	Yes
Woodcock Johnson III (Woodcock, McGrew, & Mather, 2001)	Yes	Cannot determine	Somewhat	Yes	No	No	Yes	No specific theory
Woodcock Reading Mastery Tests–Revised (Woodcock, 1998)	Yes	Yes (floor)	Yes	Yes	No	No	Yes	No specific theory

tant that a test contains a guide to the correct pronunciation of the nonwords. The majority of the tests reviewed contain a pronunciation guide for at least some of the nonwords (see Table 2).

Lastly, it is important that a test of nonword reading should be based on an experimentally validated theory of reading, allowing it to be linked to an effective and experimentally validated program of reading intervention (Rathvon, 2004). The majority of tests reviewed were linked to specific theories of reading, though some of these theories are controversial (see Table 2).

Criteria for Determining the Quality of Diagnostic Nonword Reading Tests

In addition to the characteristics listed above, there are additional requirements for a diagnostic reading test. Such a test should be capable of providing information that can lead to a specific (Tier Three) program of remediation if a child is found to have difficulties in a specific area (Hempenstall, 2009; Rathvon, 2004). To this end, it is helpful for a diagnostic test to contain an error analysis so that it is possible to see exactly where a child might need help (Kohnen, Nickels, & Castles, 2009).

Two of the 17 tests reviewed provide this kind of information: the MULTILIT Word Attack Skills Placement Test (MULTILIT, 2007), and the Word Attack subtest of the Woodcock Reading Mastery Test-Revised (WRMT-R; Woodcock, 1998). A third test, the Consortium on Reading Excellence Phonics Survey (Consortium on Reading Excellence, 2008) lists groups of sounds that are tested (such as 'short vowels', 'digraphs' and 'consonant blends'), but does not identify individual grapheme–phoneme correspondences.

Characteristics of Grapheme–Phoneme Correspondences and Nonword Stimuli

The characteristics of the grapheme–phoneme correspondences and nonword stimuli used in a test will affect how well the test is able to identify children who have difficulties with the sublexical route to reading. This is vital for diagnostic tests, where the aim is to achieve a fine-grained analysis of a student's particular skills. However, it is still relevant to Tier One and Two tests, where the aim is to identify which children are not acquiring the skills necessary to function at a level appropriate to their age.

The English orthography has at least 40 — and possibly more than 300 — grapheme–phoneme correspondences depending on how these correspondences are defined; for example, whether or not context-sensitive rules are included (Rastle & Coltheart, 1999; S. Saunders, personal communication, March 11, 2009). Ideally, the ability to apply all of these should be assessed. Yet, given the number of possible correspondences, it is difficult to conceive of a test that could include every grapheme–phoneme correspondence without being too time-consuming to administer, particularly in the case of Tier One and Two assessments.

There is considerable overlap between the grapheme–phoneme correspondences assessed by nonword reading tests. However, the manuals of the tests reviewed here do not describe the basis on which the grapheme–phoneme correspondences were chosen. It seems important to provide the rationale for these choices for various reasons. For example, in order for a test to be a truly useful measure of nonword reading skill, it should assess the grapheme–phoneme correspondences that a child is likely to need to use most often. In addition, a test that assesses only a limited range of grapheme–phoneme correspondences might underestimate or fail to reveal the true extent of a child's difficulties with decoding, while a test that assesses particularly infrequent or

difficult grapheme–phoneme correspondences might overestimate the extent of a decoding problem.

Other characteristics of nonword stimuli are also important as they can affect how easily a nonword is read (Coltheart & Leahy, 1992; Jared, 1997; Laxon, Masterson, & Coltheart, 1991; Treiman, Goswami, & Bruck, 1990). Moreover, some types of nonwords can be read with some assistance from the lexical route (Coltheart & Leahy, 1996; Treiman et al., 1990), which is not ideal when we wish to assess sublexical route functioning. These characteristics are discussed in detail below.

Frequency

Some grapheme–phoneme correspondences occur much more often than others (Coltheart et al., 2001; Fry, 2004). Children are likely to have less difficulty with frequently occurring correspondences than with those to which they have little exposure (Jackson & Coltheart, 2001). It follows then that this variable may affect how well children score on a particular test. Only one of the tests reviewed reports that grapheme–phoneme frequency was taken into account: the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999).

Consistency

According to Jackson and Coltheart (2001, p. 45), a nonword is consistent:

> *if, in all the real words that have the string's orthographic body, there is only one pronunciation of that body … Thus, the word* bean *and the pseudoword* fean *are consistent, since all the words with the body–*ean *have the rime /-in/, whereas the word* bead *and the pseudoword* fead *are inconsistent, because of the existence of words such as* head

Although consistency is usually regarded as a property of bodies (all the letters after the initial consonants; Jackson & Coltheart, 2001), grapheme–phoneme correspondences themselves can also have different levels of consistency (Voudsen, 2008). For example, the grapheme *b* is always pronounced /*b*/, meaning that its pronunciation is consistent. On the other hand, the grapheme *i* is pronounced differently in the words *pin* and *pint*, meaning that it is inconsistent. The consistency of particular word bodies has been shown to affect the ease with which nonwords can be read by children and adults (Coltheart & Leahy, 1992; Jared, 1997; Laxon et al., 1991). Therefore, it is also plausible that the consistency of individual grapheme–phoneme correspondences could also affect how easily nonwords are read (Jared, 1997). None of the tests reviewed report consistency data for the grapheme–phoneme correspondences used or for their nonword stimuli (see Table 3).

Consistency is also an issue when nonword items are created by changing letters of irregular words (e.g., *plood* derived from *flood*), because nonwords created this way will have inconsistent bodies. The Nonword Reading Test (Wren, 2003) is a good example of the problems associated with this. Items in this test were created by changing one letter of an irregular word, and irregular pronunciations of the nonwords are accepted as correct. For example, the nonword *plood* is an item on this test and it can be read to rhyme with *flood* or to rhyme with *food*. If it is read to rhyme with *flood*, it is possible that a child may have read it by analogy to the real word from which it was derived. In other words, nonwords created by changing one letter of an irregular word can be read by capitalising on the activation of the lexical route by the irregular stimuli. Thus, such nonwords are not pure measures of use of the sublexical route.

Word Position

Research has shown that the position of a grapheme–phoneme correspondence within a word can affect how accurately the correspondence is read (McCandliss, Beck, Sandak, & Perfetti, 2003). Therefore, a test aiming to assess how well a child can read individual grapheme–phoneme correspondences should ensure that each grapheme–phoneme correspondence is presented in all orthographically legal word positions, where possible. The only test reviewed that assesses individual grapheme–phoneme correspondences in multiple word positions is the WRMT-R (Woodcock, 1998).

Complexity

Complex, multiletter graphemes tend to be more difficult to read than simple, single-letter graphemes (Siegel & Faux, 1989; Snowling, 1981; Szeszulski & Manis, 1987). Both simple and complex grapheme–phoneme correspondences should be assessed. The two tests that list the grapheme–phoneme correspondences assessed — the MULTILIT Word Attack Skills Placement Test (MULTILIT, 2007) and the WRMT-R (Woodcock, 1998) — evaluate both simple and complex correspondences.

Number of Times each Correspondence is Tested

If a test aims to determine a child's knowledge of a particular grapheme–phoneme correspondence, it is insufficient to assess each grapheme–phoneme correspondence only once, as a child may not respond consistently to the grapheme–phoneme correspondences, and this information will not be revealed by a single instance of testing. In order to be certain of a child's ability to read a grapheme–phoneme correspondence, each correspondence should be assessed multiple times. Tests reviewed that assess individual grapheme–phoneme correspondences more than once are the WRMT-R (Woodcock, 1998), the MULTILIT Word Attack Skills Placement Test (MULTILIT, 2007) and the CORE Phonics Survey (Consortium on Reading Excellence, 2008; see Table 3).

Orthographic Legality

It is important that nonwords are orthographically legal. In other words, nonwords should contain letter combinations that occur in the English language, such as *brinp*, rather than letter combinations that do not occur, such as *gzanp* (Thompson, 1985; Venezky, 1970). Like adults, children are likely to have more difficulty reading orthographically illegal nonwords than orthographically legal nonwords (Grainger, Bouttevin, Truc, Bastien, & Ziegler, 2003), and there is little point in requiring children to read illegal letter combinations as they will rarely need to do this in everyday life. Seven of the tests reviewed report that the nonwords used are orthographically legal or orthographically regular in some respect, although it is not always clear how orthographic legality was established (see Table 4).

Pseudohomophones

Tests that contain items that are pseudohomophones (that is, nonwords that sound like real words when read aloud; for example, *shyne, sirf*) may overestimate a child's ability to use grapheme–phoneme correspondence rules: it has been shown that young children are able to read pseudohomophones more easily than nonwords that are not pseudohomophones (Laxon, Masterson, Gallagher, & Pay, 2002). Ten of the tests reviewed did not contain any pseudohomophones; see Table 4. However, the other seven tests reviewed contained at least one pseudohomophone, with the Dynamic Indicators of Basic Literacy Skills (DIBELS; Good & Kaminski, 2010) and the Queensland Inventory of Literacy (QUIL; Dodd et al., 1996) containing particularly high numbers of pseudohomophones.

TABLE 3

Summary of Nonword Reading Test Review: Grapheme–Phoneme Correspondences (GPCs)

Test	Reports which GPCs are assessed	Reasons for choosing particular GPCs reported	GPC frequency taken into account	GPC consistency taken into account	GPCs tested in all legal word positions	Simple/complex GPCs systematically assessed	Each GPC tested multiple times
Castles & Coltheart Reading Test 2 (Castles et al., 2009)	No	No	Not reported	Not reported	No*	No*	No*
Consortium on Reading Excellence Phonics Survey (Consortium on Reading Excellence, 2008)	Yes	No	Not reported	Not reported	No*	No*	Yes
Dynamic Indicators of Basic Literacy Skills (DIBELS), 6th edition (Good & Kaminski, 2010)	No	No	Not reported	Not reported	No*	No*	No*
The Dyslexia Screening Test – Junior (Fawcett & Nicolson, 2004)	No	No	Not reported	Not reported	No*	No*	No*
Graded Nonword Reading Test (Snowling, Stothard, & McLean, 1996)	No	No	Not reported	Not reported	No*	No*	No*
Illinois Test of Psycholinguistic Abilities, 3rd edition (Hammill, Mather, & Roberts, 2001)	No	No	Not reported	Not reported	No*	No*	No*
MULTILIT Word Attack Skills Placement Test (MULTILIT, 2007)	Yes	No	Not reported	Not reported	No	No	No
Nonword Decoding Test (Turner, 2003)	No	No	Not reported	Not reported	No*	No*	No*
Nonword Reading Test (Martin & Pratt, 2001)	No	No	Not reported	Not reported	No*	No*	No*
Nonword Reading Test (Wren, 2003)	No	No	Not reported	Not reported	No*	No*	No*
Phonological Assessment Battery (Frederickson, Frith, & Reason, 1997)	No	No	Not reported	No*	No*	No*	No*

TABLE 3 (CONTINUED)

Summary of Nonword Reading Test Review: Grapheme–Phoneme Correspondences (GPCs)

Test	Reports which GPCs are assessed	Reasons for choosing particular GPCs reported	GPC frequency taken into account	GPC consistency taken into account	GPCs tested in all legal word positions	Simple/complex GPCs systematically assessed	Each GPC tested multiple times
Queensland University Inventory of Literacy (Dodd, Holm, Oerlemans, & McCormick, 1996)	No	No	Not reported	Not reported	No*	No*	No*
Sutherland Phonological Awareness Test–Revised (Neilson, 2003)	No	No	Not reported	Not reported	No*	No*	No*
Test of Word Reading Efficiency (Torgesen, Wagner, & Rashotte, 1999)	No	No	Yes	Not reported	No*	No*	No*
Wechsler Individual Achievement Test – 2nd Australian edition (The Psychological Corporation, 2002)	No	No	Not reported	Not reported	No*	No*	No*
Woodcock Johnson III (Woodcock, McGrew, & Mather, 2001)	No	No	Not reported	Not reported	No*	No*	No*
Woodcock Reading Mastery Tests–Revised (Woodcock, 1998)	Yes	No	Not reported	Not reported	Yes	No	Yes

*Difficult to determine as the test does not report which GPCs are assessed, but does not seem to be the case.

Similarity to Real Words

It has also been shown that nonwords that are very similar to real words (for example, nonwords that have a body that is the same as a real word, such as *zame*, which shares a body with *same*) are easier to read than other nonwords, especially when the nonword shares a body with a high-frequency real word (Laxon et al., 1991; Treiman et al., 1990). This has generally been shown to affect beginning readers to a lesser extent than more experienced readers, perhaps due to the fact that experienced readers have had more exposure to the written form of words — this implies there may be some level of activation of the orthographic input lexicon (Coltheart & Leahy, 1996; Treiman et al., 1990). Thus, nonwords that are very similar to real words may not be pure measures of the sublexical route. Therefore, if possible, grapheme–phoneme correspondences should not be tested exclusively in nonwords that have bodies of high-frequency real words.

Embedded Real Words

All but one of the tests reviewed (Frederickson, Frith, & Reason's, 1997, Phonological Assessment Battery [PhAB]) contain items that have embedded real words. However, the inclusion of these items is problematic. For example, if a nonword test item contains a real word, such as *blamp* (which contains the word *lamp*), a child could sound out the single grapheme *b* and read the word *lamp* using the lexical route. This implies some ability to sound out using grapheme–phoneme correspondences, but also implies activation of the lexical route. Therefore, nonwords containing embedded real words may not be the best measure of the sublexical route, and such nonwords should be avoided as far as possible, especially in diagnostic tests of nonword reading.

Syllable/Consonant-Vowel Structure

The consonant-vowel structure of a nonword will also affect how easily it is read — for example, a nonword containing consonant clusters, such as *splomp*, is more difficult to read than a word with a simple consonant-vowel-consonant structure, such as *bip* (Snowling, 1981). Snowling (1981) hypothesised that this is due to difficulties with both segmenting the nonwords into graphemes and assigning these graphemes to phonemes. Three of the tests reviewed use the presence of consonant clusters as one way of manipulating nonword difficulty, placing nonwords without consonant clusters at the beginning of the test as 'easy' items, and items with consonant clusters at the end of the test as 'hard' items. These tests are the TOWRE (Torgesen et al., 1999), the CC2 (Castles et al., 2009) and the Graded Nonword Reading Test (Snowling, Stothard, & McLean, 1996). Because the presence of consonant clusters may affect how easily a nonword is read, it is useful for a diagnostic nonword test to assess consonant grapheme–phoneme correspondences both in isolation and as part of a cluster. None of the tests reviewed do this systematically (see Table 3).

Number of Syllables

The number of syllables in a word can also affect how easily a nonword is read, particularly for children with reading difficulties (Snowling, 1981). Generally, the greater the number of syllables, the more difficult they are to read (Duncan & Seymour, 2003; Snowling, 1981). Reading nonwords with multiple syllables is thought to be harder than reading monosyllabic nonwords because, in addition to decoding the grapheme–phoneme correspondences within the nonword, suprasegmental variables must also be taken into account, such as the assignment of syllable boundaries and stress (Duncan & Seymour, 2003; Rastle & Coltheart, 2000).

TABLE 4

Summary of Test Review: Nonwords

Test	Nonwords reported orthographically legal	Contains pseudo-homophones	Contains nonwords very similar to real words and/or embedded real words	GPCs assessed only in monosyllabic nonwords
Castles & Coltheart Reading Test 2 (Castles et al., 2009)	No	No	Yes	No
Consortium on Reading Excellence Phonics Survey (Consortium on Reading Excellence, 2008)	No	Yes	Yes	No
Dynamic Indicators of Basic Literacy Skills (DIBELS), 6th edition (Good & Kaminski, 2010)	No	Yes	Yes	Yes
The Dyslexia Screening Test–Junior (Fawcett & Nicolson, 2004)	No	No	Yes	No
Graded Nonword Reading Test (Snowling, Stothard, & McLean, 1996)	Yes	No	Yes	No
Illinois Test of Psycholinguistic Abilities, 3rd edition (Hammill, Mather, & Roberts, 2001)	Yes	No	Yes	No
MULTILIT Word Attack Skills Placement Test (MULTILIT, 2007)	Yes	No	Yes	Yes
Nonword Decoding Test (Turner, 2003)	No	Yes	Yes	No
Nonword Reading Test (Martin & Pratt, 2001)	Yes	No	Yes	No
Nonword Reading Test (Wren, 2003)	No	No	Yes	No
Phonological Assessment Battery (Frederickson, Frith, & Reason, 1997)	Yes	No	No	No
Queensland University Inventory of Literacy (Dodd, Holm, Oerlemans, & McCormick, 1996)	No	Yes	Yes	No
Sutherland Phonological Awareness Test–Revised (Neilson, 2003)	No	No	Yes	No
Test of Word Reading Efficiency (Torgesen, Wagner, & Rashotte, 1999)	No	Yes	Yes	No
Wechsler Individual Achievement Test – 2nd Australian edition (The Psychological Corporation, 2002)	Yes	Yes	Yes	No
Woodcock Johnson III (Woodcock, McGrew, & Mather, 2001)	Yes	No	Yes	No
Woodcock Reading Mastery Tests–Revised (Woodcock, 1998)	No	Yes	Yes	Yes

Many of the tests reviewed, such as the TOWRE (Torgesen et al., 1999), the PhAB (Frederickson et al., 1997) and the Graded Nonword Reading Test (Snowling et al., 1996) contain polysyllabic nonwords — these are considered the most difficult type of item (Snowling et al., 1996). In a Tier One or Tier Two test, this might not be an issue. However, in a test that aims to assess how well a child can read individual grapheme–phoneme correspondences, polysyllabic items might be misread not because a child does not know a grapheme-phoneme correspondence, but because they do not know how to assign syllable boundaries, or some other suprasegmental variable. Because these variables are not of interest in a diagnostic grapheme–phoneme correspondence test, it is beneficial if such a test contains only monosyllabic nonwords.

Summary

Various characteristics of nonword stimuli, and the grapheme–phoneme correspondences they contain, affect how easily these nonwords are read and may affect the accuracy of children's reading. These properties include the frequency and consistency of a grapheme–phoneme correspondence; the complexity of a correspondence; the position of a correspondence within a nonword; whether the correspondence is a singleton or part of a cluster; whether or not a nonword is a pseudohomophone, or shares a body with a frequent real word; whether or not a nonword contains an embedded real word; and the syllable structure of a nonword.

Nonword Tests for Assessing Progress in Acquisition of Sublexical Reading Skills

The list of criteria for nonword reading tests, test items and grapheme–phoneme correspondences is long. However, when a test is aiming to assess the general functioning of the sublexical route as compared to the typical developmental trajectory in Tier One and Tier Two assessment, it may not be necessary to take into account *all* of the factors mentioned above, and in a practical sense it may be difficult to do so.

Nevertheless, there are three tests of nonword reading available that fulfil the majority of the above criteria. These are the Martin and Pratt Nonword Reading Test (Martin & Pratt, 2001), the Woodcock Reading Mastery Test-Revised (Woodcock, 1998) and the Test of Word Reading Efficiency (TOWRE; Torgesen et al., 1999). These tests will be discussed in some detail below. For further information on all other tests reviewed, refer to Appendices B, C, D and E.

The Martin and Pratt Nonword Reading Test (Martin & Pratt, 2001) is a standardised test of nonword reading that consists of two parallel forms with 54 items in each form. It takes approximately 10 minutes to administer and is easy to both administer and score, with coloured Standard Score charts to assist with interpretation. The test has sufficient reliability and validity and is not subject to floor or ceiling effects. It also contains phonological transcriptions of correct nonword pronunciations. Items are in order of difficulty based on the performance of the standardisation sample and there is a stopping rule. The test contains reasonably recent (1996/1997) Australian norms. The test contains two parallel forms, which allow for assessment of a student's response to intervention without concerns that improved scores may be due to practice effects.

The Woodcock Reading Mastery Test-Revised (WRMT-R; Woodcock, 1998) is a standardised test of various components of reading skill, including nonword reading (the Word Attack subtest), letter and word identification and passage comprehension. It is a comprehensive test with parallel forms, and includes an error analysis (see *Nonword Tests for Diagnostic Purposes*). The Word Attack subtest contains 45 items in order of

difficulty and there is a stopping rule. The test has sufficient reliability and validity and there is a guide to the pronunciation of the nonwords. The test is easy to administer. Two disadvantages are that the test is somewhat complex to score and is subject to floor effects for the youngest age groups in the norming sample (ages 5 to 7 years and three months) where a score of zero or one falls within the normal range.

The TOWRE (Torgesen et al., 1999) is a standardised test of reading fluency with sight word and nonword subtests. Reading fluency has been identified as an important component of skilled reading because slow, laborious word decoding will impair a child's ability to comprehend a text (DEST, 2005; Hempenstall, 2009). In order to measure fluency, the subtests of the TOWRE are timed. The test is easy to administer, although it does require the use of a stopwatch, and easy to score. There is a guide to the pronunciation of the nonword items and there are no floor or ceiling effects. The test also contains alternate forms, which are an advantage when testing a child's progress over a short period of time. Although the test does not have Australian norms, the test has good reliability and validity data. The TOWRE is the only test of reading fluency available that contains nonword stimuli.

There are two other tests that fulfil most — though not all — of the above criteria and are useful for specific purposes. These tests are the Graded Nonword Reading Test (Snowling et al., 1996), and the Castles and Coltheart Reading Test 2 (CC2; Castles et al., 2009).

The Graded Nonword Reading Test (Snowling et al., 1996) is a standardised test of nonword reading. The test is very quick and easy to administer and score. It contains a clear guide to pronunciation of the nonword items. It also contains a stopping rule. However, the test is subject to floor effects for the youngest age groups in the norming sample and there is also some evidence of ceiling effects for the oldest age groups. Additionally, the test has no Australian norms. However, the test is useful as a quick screener of sublexical reading skill for ages 6 and up.

The CC2 (Castles et al., 2009) is a standardised test of reading designed to assess both the lexical and sublexical routes for reading, based on the Dual Route theory of reading. It contains regular word, irregular word and nonword items which are presented intermixed. The CC2 is the only reading test reviewed that systematically assesses regular, irregular and nonword reading skills in the one test. The test is freely available in paper and online versions from the Macquarie Online Test Interface (MOTIf; http://www.motif.org.au). The test contains 40 items of each type presented in order of difficulty. It is quick and relatively easy to administer, although when administering the paper version the examiner must keep track of how many of each type of item a child has responded to incorrectly, as the discontinuation rule is applied to each item type separately. The test has recent Australian norms. Unfortunately, the test does not contain information on reliability or validity and there is no guide to the pronunciation of the items. There is also evidence of floor effects for the two youngest age groups in the norming sample (ages 6; 0–6; 11).

Nonword Tests for Diagnostic Purposes

All the aforementioned tests of nonword reading assess a child's knowledge of grapheme–phoneme correspondence rules in relation to their peers or to a particular criterion. This is important as a first step in identifying whether a child has a problem in acquiring sublexical reading skills and is useful for deciding whether a child requires Tier Two and/or Tier Three instruction — but where does a teacher or reading specialist go from there? In order to effectively target reading instruction individually at the Tier

Three level, it may not be enough to know a child's score on a test. The specific rules a child does not know must be identified so that intervention can focus on these.

Two of the 17 tests reviewed provide this kind of information: the MULTILIT Word Attack Skills Placement Test (MULTILIT, 2007) and the Word Attack subtest of the Woodcock Reading Mastery Test-Revised (WRMT-R; Woodcock, 1998).[3]

However, although the MULTILIT Word Attack Skills Placement Test (MULTILIT, 2007) and the Word Attack subtest of the WRMT-R (Woodcock, 1998) are capable of providing information on which grapheme–phoneme correspondences a child knows, neither test fulfils all of the criteria desirable in a diagnostic test of nonword reading.

The MULTILIT Word Attack Skills Placement Test was not designed to be a pure measure of nonword reading skill — it contains both nonwords and real words and does not provide a separate score for nonword reading performance. In addition, the authors of the MULTILIT Word Attack Skills Placement Test do not report on why they chose to assess particular rules or on whether frequency and consistency of rules were taken into account. It is not clear if each rule is tested in all orthographically legal word positions and, although it is stated that the nonwords are 'phonetically regular', there is no mention of whether or not the words are orthographically legal. The test contains no pronunciation guide and some nonword test items contain embedded real words (such as *glip* and *bant*). Finally, the test is only available for use as part of the MULTILIT program of instruction and, as such, is not readily available.

The WRMT-R (Woodcock, 1998) comes close to fulfilling the majority of the criteria desirable in a diagnostic nonword reading test. However, it too does not report why particular rules were included in the test and there is no mention of whether the frequency and consistency of each correspondence rule was taken into account. Further, it is not noted whether or not a measure of orthographic legality was used when creating nonword items. Finally, some items contain embedded real words (such as *glack*, *knap* and *straced*) and some nonwords are, in fact, real words (*din* [Form H], *nigh* and *nan* [Form G]).

Summary and Conclusion

The Response to Intervention model of reading intervention requires that children be regularly assessed to ensure that they are achieving suitable progress for their age and school level (Justice, 2006). If assessment reveals that they are not achieving at an expected level, a child is provided with more intensive and increasingly individualised instruction. An important reading skill that is often impaired is the ability to decode grapheme–phoneme correspondences. If such decoding abilities are impaired, this implies difficulties with the functioning of the sublexical route to reading (Jackson & Coltheart, 2001).

We have provided an extensive list of criteria that can be used to determine the quality of a nonword reading test. At Tiers One and Two, when assessment aims to assess the general functioning of the sublexical route as compared to the typical developmental trajectory, it may not be necessary to take into account *all* of the factors mentioned, and in practice it is difficult to do so. The three nonword reading tests that fulfil the highest number of our criteria are the Martin and Pratt Nonword Reading Test (Martin & Pratt, 2001), the Word Attack subtest of the Woodcock Reading Mastery Test-Revised (Woodcock, 1998), and the Test of Word Reading Efficiency (TOWRE; Torgesen et al., 1999).

However, if a test is aiming to explore which *particular* grapheme–phoneme correspondence rules a child knows, in order to deliver a targeted program of instruction, all

of these criteria should be taken into account to ensure that the test is a pure and sensitive measure of a child's ability to use these rules. Ideally, the way the criteria were taken into account during test development should be reported in the test manual so that tests can be compared on these measures. The above review of nonword reading tests demonstrates that there is only one test that comes close to fulfilling the majority of the listed criteria — the Woodcock Reading Mastery Test (Woodcock, 1998). However, it appears that there is no diagnostic nonword reading test in existence that takes into account all of the criteria mentioned. For this reason, we are developing a Diagnostic Reading Test for Nonwords (DiRT-Nonwords). This test is currently in the process of being normed and a preliminary version is available from the Macquarie Online Test Interface (MOTIf; http://www.motif.org.au).

Notes

1   Note that RTI intervention may consist of more than three tiers, each tier more intensive than the last, with assessment at each tier level determining whether or not a student needs to move on to a further tier. However, for convenience, only three tiers are discussed here.

2   Note that the sublexical route to reading has three components: parsing (breaking words down into graphemes — letters or letter groups such as 'CH', 'OA'), translating graphemes to phonemes (sounds), and blending these phonemes together to pronounce a word (Jackson & Coltheart, 2001; Joubert & Lecours, 2000). This article is concerned with the second component (grapheme–phoneme correspondences); however, it is also possible for a child to have difficulties with the other components of the sublexical route (Jackson & Coltheart, 2001; Joubert & Lecours, 2000).

3   A third test, the Consortium on Reading Excellence (CORE) Phonics Survey (Consortium on Reading Excellence, 2008) lists groups of sounds that are tested (such as 'short vowels', 'digraphs' and 'consonant blends'), but does not identify individual grapheme–phoneme correspondences.

Acknowledgments

References

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.

Australian Council for Educational Research Website. (2009). Retrieved April 1, 2009, from http://www.acer.edu.au/

Berninger, V.W. (2001). *Process assessment of the learner: Test battery for reading and writing.* San Antonio, TX: The Psychological Corporation.

Buros Institute of Mental Measurements. (2009). *Test reviews online.* Retrieved March 31, 2009, from http://buros.unl.edu/buros/jsp/search.jsp

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition, 47,* 149–180.

Castles, A., Coltheart, M., Larsen, L., Jones, P., Saunders, S., & McArthur, G. (2009). Assessing the basic components of reading: A revision of the Castles and Coltheart test with new norms. *Australian Journal of Learning Difficulties, 14,* 67–88. doi:10.1080/19404150902783435

Catts, H.W., Petscher, Y., Schatschneider, C., Bridges, M.S., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on early identification of reading disabilities. *Journal of Learning Disabilities*, *42*, 163–176. doi:10.1177/0022219408326219

Coltheart, V., & Leahy, J. (1992). Children's and adult's reading of nonwords: Effects of regularity and consistency. *Journal of Experimental Psychology*, *18*, 718–729.

Coltheart, V., & Leahy, J. (1996). Procedures used by beginning and skilled readers to read unfamiliar letter strings. *Australian Journal of Psychology*, *48*, 124–129. doi:10.1080/00049539608259518

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A Dual Route Cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204–256. doi:10.1037/0033–295X.108.1.204

Consortium on Reading Excellence. (2008). *Assessing reading: Multiple measures for Kindergarten through twelfth grade* (2nd ed.). Novato, CA: Arena Press.

Coyne, M.D., & Harn, B.A. (2006). Promoting beginning reading success through meaningful assessment of early literacy skills. *Psychology in the Schools*, *43*, 33–43. doi:10.1002/pits.20127

Department of Education, Science and Training. (2005). *Teaching reading: Report of the National Inquiry into the Teaching of Literacy*. Retrieved March 31, 2009, from http://www.dest.gov.au/nitl/

Dodd, B., Holm, A., Oerlemans, M., & McCormick, M. (1996). *Queensland University Inventory of Literacy*. St Lucia, Queensland: Department of Speech Pathology & Audiology, University of Queensland.

Duncan, L.G., & Seymour, P.H.K. (2003). How do children read multisyllabic words? Some preliminary observations. *Journal of Research in Reading*, *26*, 101–120. doi:10.1111/1467–9817.00190

EBSCOhost Tests in Print Database. (2009). Retrieved March 31, 2009, from the Macquarie University Library website, http://www.library.mq.edu.au

Fawcett, A.J., & Nicolson, R.I. (2004). *The Dyslexia Screening Test-Junior*. London, UK: Harcourt Assessment.

Frederickson, N., Frith, U., & Reason, R. (1997). *Phonological Assessment Battery*. London, UK: nferNelson.

Fry, E. (2004). Phonics: A large phoneme-grapheme frequency count revised. *Journal of Literacy Research*, *36*, 85–98.

Fuchs, D., Mock, D., Morgan, P.L., & Young, C.L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research & Practice*, *18*, 157–171. doi:10.1111/1540–5826.00072

Good, R.H., III, & Kaminski, R.A. (Eds.). (2010). *Dynamic indicators of basic literacy skills* (6th ed.). Eugene, OR: Dynamic Measurement Group. Retrieved 10 July, 2010, from http://www.dibels.org

Grainger, J., Bouttevin, S., Truc, C., Bastien, M., & Ziegler, J. (2003). Word superiority, pseudoword superiority, and learning to read: A comparison of dyslexic and normal readers. *Brain & Language*, *87*, 432–440. doi:10.1016/S0093–934X(03)00145–7

Hammill, D.D., Mather, N., & Roberts, R. (2001). *Illinois Test of Psycholinguistic Abilities* (3rd ed.). Austin, TX: Pro-Ed.

Harrison, G.L. (2005). A review of the Early Reading Diagnostic Assessment, Second Edition (ERDA-2). *Canadian Journal of School Psychology*, *20*, 111–116. doi:10.1177/0829573506295459

Hempenstall, K. (2009). Research-driven reading assessment: Drilling to the core. *Australian Journal of Learning Difficulties*, *14*, 17–52. doi:10.1080/19404150902783419

Hosp, J.L., & Ardoin, S.P. (2008). Assessment for instructional planning. *Assessment for Effective Intervention*, *33*, 69–77. doi:10.1177/1534508407311428

Jackson, N.E., & Coltheart, M. (2001). *Routes to reading success and failure: Toward an integrated cognitive psychology of atypical reading*. New York, NY: Psychology Press.

Jared, D. (1997). Spelling–sound consistency affects the naming of high-frequency words. *Journal of Memory and Language*, *36*, 505–529. doi:10.1006/jmla.1997.2496

Joubert, S.A., & Lecours, A.R. (2000). The role of sublexical graphemic processing in reading. *Brain & Language*, *72*, 1–13. doi:10.1006/brln.1999.2279

Justice, L.M. (2006). Evidence-based practice, response to intervention, and the prevention of reading difficulties. *Language, Speech, and Hearing Services in Schools*, *37*, 284–297. doi:10.1044/0161–1461(2006/033)

Kame'enui, E.J., Fuchs, L., Francis, D.J., Good, R., III, O'Connor, R.E., Simmons, D.C., & Torgesen, J.K. (2006). The adequacy of tools for assessing reading competence: A framework and review. *Educational Researcher*, *35*, 3–11. doi:10.3102/0013189X035004003

Knutson, J.S., Simmons, D.C., Good, R., III, & McDonagh, S.H. (2004). Specially designed assessment and instruction for children who have not responded adequately to reading intervention. *Assessment for Effective Intervention*, *29*(4), 47–58. doi:10.1177/073724770402900407

Kohnen, S., Nickels, L., & Castles, A. (2009). Assessing spelling skills and strategies: A critique of available resources. *Australian Journal of Learning Difficulties*, *14*, 113–150. doi:10.1080/19404150902783450

Laxon, V., Masterson, J., & Coltheart, V. (1991). Some bodies are easier to read: The effect of consistency and regularity on children's reading. *The Quarterly Journal of Experimental Psychology*, *43A*, 793–824. doi:10.1080/14640749108400958

Laxon, V., Masterson, J., Gallagher, A., & Pay, J. (2002). Children's reading of words, pseudohomophones, and other nonwords. *The Quarterly Journal of Experimental Psychology*, *55A*, 543–565. doi:10.1080/02724980143000479

Mahdavi, J.M., & Sawyer, D. (2004). *Review of Early Reading Success Indicator*. In the Seventeenth Mental Measurements Yearbook. Retrieved August 14, 2010, from EBSCO Mental Measurements Yearbook Database through the Macquarie University Library website, http://www.lib.mq.edu.au

Martin, F., & Pratt, C. (2001). *Nonword Reading Test.* Melbourne, Australia: ACER.

McCandliss, B., Beck, I.L., Sandak, R., & Perfetti, C. (2003). Focusing attention on decoding for children with poor reading skills: Design and preliminary tests of the word building intervention. *Scientific Studies of Reading*, *7*, 75–104.

MULTILIT. (2007). *Multilit Word Attack Skills Manual.* Sydney, Australia: MULTILIT.

National Reading Panel. (2000). *A report of the national reading panel: Teaching children to read.* Washington, DC: National Institute of Child Health and Human Development.

Neilson, R. (2003). *Sutherland Phonological Awareness Test-Revised.* Jamberoo, Australia: Roslyn Neilson.

Ovid Mental Measurements Yearbook. (2009). Retrieved March 31, 2009, and April 1, 2009, from the Macquarie University Library website, http://www.library.mq.edu.au

Peterson, L.S., Martinez, A., & Turner, T.L. (2010). Test review: Process Assessment of the Learner-Second Edition. *Journal of Psychoeducational Assessment*, *28*, 80–86. doi:10.1177/0734282909346717

Rack, J.P., Snowling, M.J., & Olson, R.K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly*, *27*, 29–53.

Rastle, K., & Coltheart, M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 482–503.

Rastle, K., & Coltheart, M. (2000). Lexical and nonlexical print-to-sound translation of disyllabic words and nonwords. *Journal of Memory and Language*, *42*, 342–364. doi:10.1006/jmla.1999.2687

Rathvon, N. (2004). *Early reading assessment: A practitioner's handbook.* New York, NY: Guilford Press.

Siegel, L.S., & Faux, D. (1989). Acquisition of certain grapheme–phoneme correspondences in normally achieving and disabled readers. *Reading and Writing*, *1*, 37–52. doi:10.1007/BF00178836

Snowling, M.J. (1981). Phonemic deficits in developmental dyslexia. *Psychological Research*, *43*, 219–234. doi:10.1007/BF00309831

Snowling, M.J., Stothard, S.E., & McLean, J. (1996). *Graded Nonword Reading Test.* Bury St Edmonds, UK: Thames Valley Test Company.

Szeszulski, P.A., & Manis, F.R. (1987). A comparison of word recognition processes in dyslexic and normal readers at two reading-age levels. *Journal of Experimental Child Psychology*, *44*, 364–376.

The Psychological Corporation. (2002). *Wechsler Individual Achievement Test* (2nd Australian ed.). San Antonio, TX: Author.

The Psychological Corporation. (2003). *Early Reading Diagnostic Assessment* (2nd ed.). San Antonio, TX: Harcourt.

Thompson, G.B. (1985). Orthographic structures: Grapheme patterns in child and adult texts. *Journal of Research in Reading*, *8*, 32–44. doi:10.1111/j.1467–9817.1985.tb00271.x

Torgesen, J.K., Wagner, R.K., & Rashotte, C.A. (1999). *Test of Word Reading Efficiency*. Austin, TX: Pro-Ed.

Treiman, R., Goswami, U., & Bruck, M. (1990). Not all nonwords are alike: Implications for reading development and theory. *Memory & Cognition*, *18*, 559–567. doi:10.3758/BF03197098

Turner, M. (2003). *The Nonword Decoding Test*. Retrieved March 31, 2009, from http://www.tes.co.uk/teaching-resource/The-Nonword-Decoding-Test-6049038/

Venezky, R.L. (1970). *The structure of English orthography*. The Hague: Mouton.

Voudsen, J.I. (2008). Units of English spelling-to-sound mapping: A rational approach to reading instruction. *Applied Cognitive Psychology*, *22*, 247–272. doi:10.1002/acp.1371

Westwood, P. (2009). Arguing the case for a simple view of literacy assessment. *Australian Journal of Learning Difficulties*, *14*, 3–15. doi:10.1080/19404150902783401

Woodcock, R.W. (1998). *Woodcock Reading Mastery Tests-Revised*. Circle Pines, MN: American Guidance Service.

Woodcock, R.W., McGrew, K.S., & Mather, N. (2001). *Woodcock Johnson III*. Itasca, IL: Riverside.

Wren, S. (2003). The Nonword Reading Test. Derived from M.J. Snowling, J. Stackhouse & J.P. Rack. (1986). Phonological dyslexia and dysgraphia: A developmental analysis. *Cognitive Neuropsychology*, 3, 309–339. Retrieved March 31, 2009, from http://www.balancedreading.com

Appendix A
Nonword Tests not Reviewed in This Article

Test name	Reason not reviewed
Fox in a Box: An Adventure in Literacy (CTB McGraw-Hill, 2002)	U.S. test, not available in Australia
The Phonological Awareness Test 2 (Robertson & Salter, 2007)	U.S. test, not available in Australia
Early Reading Diagnostic Assessment-Second Edition (The Psychological Corporation, 2003)[a]	Pseudoword decoding subtest derived from the WIAT-II (The Psychological Corporation, 2002)
Early Reading Success Indicator (ERSI; Pearson, 2009)[b]	Pseudoword decoding subtest derived from the WIAT-II (The Psychological Corporation, 2002)
Nonword Reading Test (Crumpler & McCarty, 2004)	UK test, not available in Australia
Process Assessment in the Learner (PAL) Test Battery for Reading and Writing-2nd edition (Berninger, 2001)[c]	Pseudoword decoding subtest derived from the WIAT-II (The Psychological Corporation, 2002)

Note: [a] For a review of the ERDA-2, the reader is referred to Harrison (2005). [b] For a review of the ERSI, the reader is referred to Mahdavi & Sawyer (2004). [c] For a review of the PAL-2, the reader is referred to Peterson, Martinez, & Turner (2010).

Appendix B

Review of Nonword Reading Tests: Overview

Test	Test description	Is the test based on a particular theory of reading acquisition?	What norms are there?	Information on normal error patterns?	Are there ceiling/floor effects?	Additional information
Castles & Coltheart Reading Test 2 (Castles et al., 2009)	Standardised test of reading with regular word, irregular word and nonword subtests	Yes, the dual route model	2008 Australian norms for ages 6–12 in 6-month intervals	No	Yes, floor effect for youngest age groups (6–6;5, 6;6–6;11)	Test contains irregular words, nonwords and regular words, presented in a mixed fashion
Consortium on Reading Excellence Phonics Survey (Consortium on Reading Excellence, 2008)	Criterion-based test of phonics and phonics-related skills	Research-based, but no specific theory reported	No norms – criterion-based	No	It is a test of mastery so a ceiling effect is expected, with all children eventually being able to read all items correctly	Subtest contains both real and nonwords
Dynamic Indicators of Basic Literacy Skills (DIBELS), 6th edition (Good & Kaminski, 2010)	Fluency-based measure of various reading component skills	Based on evidence from National Reading Panel (2000) and other research outlining the core skills underlying reading	Criterion-based with benchmark scores currently being gathered (available from August 2010)	No	Ceiling effect is expected – children are expected to reach certain benchmark goals	Children receive marks for number of letters correctly sounded out as well as for whole words read correctly
The Dyslexia Screening Test – Junior (Fawcett & Nicolson, 2004)	Standardised dyslexia screening measure	Based on a controversial theory which attributes dyslexia to cerebellar abnormality (Nicolson, Fawcett, & Dean, 2001)	1995–2002 UK norms for ages 6;6–11;5 in 12-month intervals	No	No	Test has practice passage. Test is intended as a screening measure. Test has measures of clumsiness and balance (children are pushed using a 'balance tester')
Graded Nonword Reading Test (Snowling, Stothard, & McLean, 1996)	Standardised test of nonword reading	Based on evidence that children with dyslexia have specific deficits in nonword reading (Rack, Snowling, & Olson, 1992)	1990–1993 UK norms for chronological ages 5–11 in 6-month intervals, and reading ages	No	Yes – floor effect for youngest age groups (5;0–5;11 and 6;11)	Test has practice items

(continued)

Appendix B (CONTINUED)

Review of Nonword Reading Tests: Overview

Test	Test description	Is the test based on a particular theory of reading acquisition?	What norms are there?	Information on normal error patterns?	Are there ceiling/floor effects?	Additional information
Illinois Test of Psycholinguistic Abilities, 3rd edition (Hammill, Mather, & Roberts, 2001)	Standardised test of psycholinguistic processing	No specific theory of reading acquisition, but based on a modified version of Osgood's (1957) theory of psycholinguistic processing	1999/2000 US norms in three-month intervals for ages 6;6–12;11 and 12-month intervals until 12;11	No	Floor effect for ages 6;6–7;2; ceiling effect for ages 8;3–12;11 (score of 24/25 falls within normal range)	Nonwords presented as the names of make-believe creatures. Discrepancy scores between real word and pseudoword reading can be calculated
MULTILIT Word Attack Skills Placement Test (MULTILIT, 2007)	Placement test designed to place students within the MULTILIT program of instruction	Dual route theory	No norms	No	Difficult to determine as no norms	Not intended as a diagnostic test separate from the MULTILIT program of instruction
Nonword Decoding Test (Turner, 2003)	Partially standardised test of nonword reading freely available online	No specific theory of reading acquisition, but based on the notion that phonological processing is the core deficit in dyslexia (Rack, 1992)	1994 UK norms – reading age equivalents given for particular raw scores	No	Difficult to determine due to lack of information about scores and norms	Not designed for obtaining scores, but designed to probe decoding technique
Nonword Reading Test (Martin & Pratt, 2001)	Standardised test of nonword reading	Based on research showing that 'efficient phonological recoding ability is a prerequisite for efficient reading'	1996/1997 Australian norms for ages 6–16 in 6-month intervals	No	No	Test has alternate forms
Nonword Reading Test (Wren, 2003)	Partially standardised test of nonwords freely available online. The test was taken from Snowling, Stackhouse, and Rack's (1986) article	Based on Frith's (1985) model of three phases in the acquisition of reading – logographic, alphabetic and orthographic	1986 norms involving 12 7-year-olds and 12 children whose reading ages were between 10 and 11	Some info on the error patterns made by people with dyslexia in Snowling et al.'s 1986 study	Difficult to determine due to lack of information about scores and norms	N/A

(continued over)

Appendix B (CONTINUED)
Review of Nonword Reading Tests: Overview

Test	Test description	Is the test based on a particular theory of reading acquisition?	What norms are there?	Information on normal error patterns?	Are there ceiling/floor effects?	Additional information
Phonological Assessment Battery (Frederickson, Frith, & Reason, 1997)	Standardised test of phonological awareness	Based on Frith's (1995) framework of components of phonological processing (PA) and other evidence relating PA to reading	1996 norms for ages 6–14;11 in 6-month intervals	No	Floor effect for age 6;0–6;5; Ceiling effect after age 8;11, (score of 19/20 falls within normal range)	Test has practice items
Queensland University Inventory of Literacy (Dodd, Holm, Oerlemans, & McCormick, 1996)	Standardised test of phonological awareness	Based on evidence that phonological awareness is linked to measures of reading and spelling	1994 grade-based norms	No	No floor effect, but ceiling effect for children in Grade 7 – score of 24/24 is within normal range	N/A
Sutherland Phonological Awareness Test – Revised (Neilson, 2003)	Standardised test of phonological awareness	Based on research which supports a connection between phonological awareness and reading/spelling skills	2003 Australian norms based on years of schooling	Some information on 'patterns of weakness'	Difficult to calculate as no percentile rank or standardised score information for individual subtest, but appears to be a floor effect	Test has alternate forms. Analysis of nonword spelling available, but not of nonword reading
Test of Word Reading Efficiency (Torgesen, Wagner, & Rashotte, 1999)	Standardised test of reading fluency with sight word and nonword subtests	Based on Ehri's (1998) model of 5 strategies used in reading and Gough's (1996) 'simple view' of reading	1999 norms for ages 6–25, 6-month intervals for ages 6 to 8; 12-month intervals for ages 8 to 24; 5-month intervals for grade	No	Yes – floor effect for age 6;0–6;11 and grades 1;0–2;11, ceiling effect for ages 17;0–24;11 and grades 11;0–12;9	Test of reading efficiency – subtests are timed. Has alternate forms.
Wechsler Individual Achievement Test – 2nd Australian edition (The Psychological Corporation, 2002)	Standardised test of academic achievement	Based on findings from 'Report of the National Reading Panel: Teaching Children to Read' (National Institute of Child Health and Human Development, 2000)	2004–2005 Australian norms for ages 6;0–19;11 in 3-month intervals and grade-based norms	No	Yes – floor effect for 6;0–6;7 and ceiling effect for ages 15;0–19;11	Test has practice items. Reliability and validity info from US sample

Appendix B (CONTINUED)

Review of Nonword Reading Tests: Overview

Test	Test description	Is the test based on a particular theory of reading acquisition?	What norms are there?	Information on normal error patterns?	Are there ceiling/floor effects?	Additional information
Woodcock Johnson III (Woodcock, McGrew, & Mather, 2001)	Standardised test of cognitive abilities	No specific theory of reading acquisition, but battery based on Cattell-Horn-Carroll factor theory of cognitive abilities	1996–1999 norms for ages 6–40+	No	Cannot determine as necessary information only available on Compuscore program for individual scores	Alternate forms available. Word Attack subtest is not part of the standard battery – part of the 'extended' battery
Woodcock Reading Mastery Tests – Revised (Woodcock, 1998)	Standardised test of various reading component skills	No specific theory – based on notion that encountering nonsense words simulates a person encountering an unknown word	1995–1996 norms for ages 5–22; one-month intervals up to age 19	No, but an error pattern analysis can be completed from a person's results	Floor effects for ages 5;0–7;3	Test has alternate forms and 2 practice items. Subtest not administered if child's score on the Word Identification subtest is 0 or 1

Appendix C (i)

Review of Nonword Reading Tests: Administration Information

Key to difficulty ratings	Ease of administration
D1 (Easy)	A test is 'easy' to administer when it does not have any special instructions beyond what you would expect from a standardised test, and when it requires no extra equipment
D2 (Slightly difficult)	A test is 'slightly difficult' to administer when a complicating factor applies, such as one of the following:
	(a) There are special instructions (e.g., you have to establish a basal/ceiling)
	(b) You need to use any extra equipment (e.g., a stopwatch)
	(c) Specialised knowledge is required (e.g., knowledge of a particular phonetic transcription system)
D3 (Difficult)	A test is 'difficult' to administer when two or more of the above conditions apply
	Ease of scoring
D1 (Easy)	A test is 'easy' to score if it involves basic ticking boxes/circling responses, basic adding up and has standard/z-score score tables that are easy to use
D2 (Slightly difficult)	A test is 'slightly difficult' to score if a complicating factor applies, such as one of the following:
	(a) It involves any calculations more difficult than basic adding
	(b) Standard score/percentile/age equivalent tables are difficult to use
	(c) There is no guidance as to how the answers should be correctly pronounced
	(d) Scores are difficult to interpret meaningfully due to lack of norms or lack of reporting of standard scores, percentiles, etc.
D3 (Difficult)	A test is 'difficult' to score if two or more of the above apply

Appendix C
Review of Nonword Reading Tests: Overview

Test	How many items are there?	How long does the test take to administer?	How difficult is the test to administer?	How difficult is the test to score?	Is there a guide to pronunciation?	Reliability/validity	Is there a stopping rule?
Castles & Coltheart Reading Test 2 (Castles et al., 2009)	40	Approximately 10 mins	D2a (Slightly difficult – need to apply stopping rule to each item type separately)	D2a (Slightly difficult as tester must score 3 different item types separately)	No	No reliability or validity data provided	Yes – after 5 consecutive errors on a particular item type, that item type is discontinued
Consortium on Reading Excellence Phonics Survey (Consortium on Reading Excellence, 2008)	35 monosyllabic, 8 bisyllabic	Whole test 10–15 mins	D1 (Easy)	D2d (Easy to score but difficult to interpret as no norms)	No, but they do indicate that stress can be placed on either syllable	No reliability or validity data provided	Yes – nonword items are not administered if student cannot read 2 or more real words within a particular item set
Dynamic Indicators of Basic Literacy Skills (DIBELS), 6th edition (Good & Kaminski, 2010)	50 per progress monitoring/ benchmark assessment	Approximately 2 mins	D2b (Slightly difficult – requires use of a stopwatch)	D3ad (Difficult – scoring instructions are complicated, information on percentile ranks, etc., difficult to obtain outside US)	Yes	Data available in the DIBELS Next Technical Manual	Yes, if student gets no sounds correct in the first row of testing items. Test is timed (1 min).
The Dyslexia Screening Test – Junior (Fawcett & Nicolson, 2004)	Ages 6;8–8;5: 38, Ages 8;6–10;5: 46, Ages 10;6–11;5: 52	Approximately 4 mins	D2b (Slightly difficult – requires use of a stopwatch)	D3ac (Difficult as need to compute basic score · child gets 2 points if correct, 1 for 'close try', bonus points or penalty score if under or over time	No	Some data on reliability and validity provided	Yes, after 5 consecutive errors, and test is timed (3 mins per passage)

(continued over)

Appendix C (CONTINUED)
Review of Nonword Reading Tests: Overview

Test	How many items are there?	How long does the test take to administer?	How difficult is the test to administer?	How difficult is the test to score?	Is there a guide to pronunciation?	Reliability/validity	Is there a stopping rule?
Graded Nonword Reading Test (Snowling, Stothard, & McLean, 1996)	20	Approximately 5 mins	D1 (Easy)	D1 (Easy)	Yes, words are phonetically transcribed with transcription key for those not familiar with transcription symbols	Some data on reliability and validity provided	Yes, after 6 consecutive errors
Illinois Test of Psycholinguistic Abilities – 3rd edition (Hammill, Mather, & Roberts, 2001)	25	Not reported for individual subtest, approximately 5–10 mins	D2a (Slightly difficult as basals and ceilings must be established)	D1 (Easy)	No	Some data on reliability and validity provided	Yes, after 3 consecutive items incorrect
MULTILIT Word Attack Skills Placement Test (MULTILIT, 2007)	133	Not reported, approximately 5–10 mins	D1 (Easy)	D2d (Easy to score but difficult to interpret as no norms)	No	No reliability or validity data provided	Yes – a level is failed after 2 items incorrect, testing stops after 2 failed levels
Nonword Decoding Test (Turner, 2003)	39	Not reported, approximately 5–10 mins	D1 (Easy)	D2d (Easy to score but difficult to interpret as no norms)	For some items	Minimal information on reliability and validity provided	Yes, after 6–10 items failed

Appendix C (CONTINUED)
Review of Nonword Reading Tests: Overview

Test	How many items are there?	How long does the test take to administer?	How difficult is the test to administer?	How difficult is the test to score?	Is there a guide to pronunciation?	Reliability/validity	Is there a stopping rule?
Nonword Reading Test (Martin & Pratt, 2001)	108 (54 per form)	~10 mins	D1 (Easy)	D1 (Easy) – coloured standard score charts assist with interpretation	Yes, phonetic transcription with key	Good data on reliability and validity provided	Yes, after 8 consecutive items incorrect
Nonword Reading Test (Wren, 2003)	31	Not reported, approximately 5–10 mins	D1 (Easy)	D2d (Easy to score but difficult to interpret as no norms)	No – states that answers can be produced regularly or irregularly	No reliability or validity data provided	No
Phonological Assessment Battery (Frederickson, Frith, & Reason, 1997)	20	Approximately 5 mins	D1 (Easy)	D1 (Easy)	Yes – real words containing appropriate sounds	Good data on reliability and validity provided	Yes – items on card 3 are not administered if 6 or more errors are made on card 2
Queensland University Inventory of Literacy (Dodd, Holm, Oerlemans, & McCormick, 1996)	24	Not reported for individual subtest, approxmately 5 mins	D1 (Easy)	D2c (Easy to score, but slightly difficult to interpret as information about correct pronunciation is only given for some items)	No	Some data on reliability and validity provided	Yes, after 4 consecutive errors

Appendix C (CONTINUED)
Review of Nonword Reading Tests: Overview

Test	How many items are there?	How long does the test take to administer?	How difficult is the test to administer?	How difficult is the test to score?	Is there a guide to pronunciation?	Reliability/validity	Is there a stopping rule?
Sutherland Phonological Awareness Test – Revised (Neilson, 2003)	8	Not reported for individual subtest, approximately 5 mins	D2a (Slightly difficult as instructions can be ambiguous)	D2d (Easy to score but difficult to interpret due to the way subtest norms are presented)	Yes, for some items	Minimal information on reliability and validity provided	Yes, if child has 'no strategies for attempting the first two items, or seems stressed'
Test of Word Reading Efficiency (Torgesen, Wagner, & Rashotte, 1999)	124 (62 per form)	5–8 mins	D2b (Slightly difficult – requires use of a stopwatch)	D1 (Easy)	Yes – real words containing appropriate sounds and phonetic transcription	Good data on reliability and validity provided	No, but students have only 45 secs to respond
Wechsler Individual Achievement Test – 2nd Australian edition (The Psychological Corporation, 2002)	55	Not reported for individual subtest, approximately 5–10 mins	D1 (Easy)	D1 (Easy)	Yes – on a CD	Good data on reliability and validity provided	Yes, after 7 consecutive errors
Woodcock Johnson III (Woodcock, McGrew, & Mather, 2001)	32	5–10 mins	D2a (Slightly difficult – need to establish basals and ceilings)	D1 (Easy)	Yes, for most items	Good data on reliability and validity provided	Yes, after 6 consecutive errors
Woodcock Reading Mastery Tests – Revised (Woodcock, 1998)	45	Not reported for individual subtest, approximately 5–10 mins	D1 (Easy)	D2b (Easy to calculate raw score but complex system of norm-referenced scores makes it difficult to interpret)	Yes	Good data on reliability and validity provided	Yes, after 6 consecutive errors

Appendix D

Review of Nonword Reading Tests: Grapheme–Phoneme Correspondences

Test	Are individual GPC rules explicitly tested?	How is knowledge of GPC rules tested?	Which GPC rules are tested?	How many times is each GPC tested?	How was the inclusion/exclusion of a rule decided upon?
Castles & Colheart Reading Test 2 (Castles et al., 2009)	No	In 1–2 syllable nonwords (peng, shoathe, floatchtwail)	Not reported	Not reported	Not reported
Consortium on Reading Excellence Phonics Survey (Consortium on Reading Excellence, 2008)	Groups/categories of rules assessed rather than individual rules	In 1 and 2 syllable nonwords (dit, pem, dilt, qued, dran, timbut, gromu, fauntoon)	Short/long vowels, digraphs, consonant blends, r-controlled and variant vowels, low frequency and consonant spellings, open/closed syllables, silent -e, consonant -le, 'vowel team'	Monosyllabic words – 5, bisyllabic words – 2	Not reported
Dynamic Indicators of Basic Literacy Skills (DIBELS), 6th edition (Good & Kaminski, 2010)	No	In monosyllabic CV and CVC nonwords (ub, tej, bal, lof)	Not reported	Not reported	Not reported
The Dyslexia Screening Test – Junior (Fawcett & Nicolson, 2004)	No	In 1–3 syllable nonwords (mib, gruny, widdyfer) embedded in reading passages	Not reported	Not reported	Not reported
Graded Nonword Reading Test (Snowling, Stothard, & McLean, 1996)	No	In 1–2 syllable nonwords (gromp, hinsink)	Not reported	Not reported	Not reported
Illinois Test of Psycholinguistic Abilities, 3rd edition (Hammill, Mather, & Roberts, 2001)	No	In 1–4 syllable nonwords (lum, blutter, hicabaffer)	Not reported	Not reported	Not reported
MULTILIT Word Attack Skills Placement Test (MULTILIT, 2007)	Yes	Single sounds, 1-syllable nonwords (fhus, moil, sprang) and regular words	Single vowel and consonant sounds, consonant digraphs (sh, ch, ng, wh, kn, ck, th); final -e rule; ee, ea, ai, ay, oo, ew, ue, oa, ow, oe, ou, ow, ir, er, ur, oy, oi, ie, igh, y, ar, au, aw	Different for each rule, usually 1–3 times per rule	Not reported

Appendix D (CONTINUED)

Review of Nonword Reading Tests: Grapheme–Phoneme Correspondences

Test	Are individual GPC rules explicitly tested?	How is knowledge of GPC rules tested?	Which GPC rules are tested?	How many times is each GPC tested?	How was the inclusion/ exclusion of a rule decided upon?
Nonword Decoding Test (Turner, 2003)	No	In 1–5 syllable nonwords (bos, freggy, cardonite, concenated, apprixengilate)	Not reported	Not reported	Not reported
Nonword Reading Test (Martin & Pratt, 2001)	No	In 1–3 syllable nonwords (vot, prilt, ficket, conarcy)	Not reported	Not reported	Not reported
Nonword Reading Test (Wren, 2003)	No	In 1–2 syllable nonwords (wamp, hausage, polonel) created by changing one letter of each word in a list of irregular words created by Snowling, Stackhouse, and Rack (1986)	Not reported	Not reported	Not reported
Phonological Assessment Battery (Frederickson, Frith, & Reason, 1997)	No	In 1–2 syllable nonwords (tib, chog, ropsatch)	Not reported	Not reported	Not reported
Queensland University Inventory of Literacy (Dodd, Holm, Oerlemans, & McCormick, 1996)	No	In 1–3 syllable nonwords (slet, querdly, didderent)	Not reported	Not reported	Not reported
Sutherland Phonological Awareness Test – Revised (Neilson, 2003)	No	In 1–3 syllable nonwords (mup, trom, fouse, mespuntal)	Not reported	Not reported	Not reported
Test of Word Reading Efficiency (Torgesen, Wagner, & Rashotte, 1999)	No	In 1–3 syllable nonwords	Not reported	Not reported	Not reported

(continued)

Appendix D (CONTINUED)

Review of Nonword Reading Tests: Grapheme–Phoneme Correspondences

Test	Are individual GPC rules explicitly tested?	How is knowledge of GPC rules tested?	Which GPC rules are tested?	How many times is each GPC tested?	How was the inclusion/ exclusion of a rule decided upon?
Wechsler Individual Achievement Test – 2nd Australian edition (The Psychological Corporation, 2002)	No	In 1–3 syllable nonwords (heb, pragment, tomingly)	Not reported	Not reported	Not reported
Woodcock Johnson III (Woodcock, McGrew, & Mather, 2001)	No	In single letters and 1–5 syllable nonwords (ib, floxy, phintober, querpostonious)	Not reported	Not reported	Not reported
Woodcock Reading Mastery Tests – Revised (Woodcock, 1998)	Yes	In 1–4 syllable nonwords (ip, rejune, darlanker, ceisminadolt)	-b, -b-, ch-, d-, -ed, -f, ph-, -g, h-, wh-, -dge, -ck, -k, qu-, m-, -m, -mb, -n, kn-, pn-, -p, r-, rh-, wr-, s-, -ss, c-, sh-, -t, t-, -th, v-, -v, w-, -x, y-, z -, -ft, gl-, -nd, dr-, -l's, -nk, sl-, -n't, br-, -cked, -lt, thr-, spr-, squ-, -a-, - e-, e-, i-, -i-, o-, -o-, -u-, -ay, -ay-, -ee, -ea, -ie-, -oa-, -o-e, -aw, -au-, - oy, -ow, -ur, -ir	1–6 (different for each rule)	Not reported

Appendix E
Review of Nonword Reading Tests: Nonwords

Test	Is GPC/nonword consistency taken into account?	Is rule frequency taken into account?	Are items orthographically legal?	Do the items contain parts that are real words?	Is each GPC tested in all orthographically legal word positions?	Do the items include pseudo-homophones?
Castles & Coltheart Reading Test 2 (Castles et al., 2009)	Not reported	Not reported	Not reported	Yes – blick, spatch, crat, framp	Not reported	No
Consortium on Reading Excellence Phonics Survey (Consortium on Reading Excellence, 2008)	Not reported	Not reported	Not reported	Yes – nask, cang, dran, chid, phid, carn, faim, soat, fout, bice, knod, timbut, podated, pentate	Not reported	Yes – faim, knod, loe, beap
Dynamic Indicators of Basic Literacy Skills (DIBELS), 6th edition (Good & Kaminski, 2010)	Not reported	Not reported	Reports items are 'phonetically regular'	Yes – nan, wan	Not reported	Yes – sez, bil, kup, sak, roc, kik, pik, yel, luv, eg
The Dyslexia Screening Test – Junior (Fawcett & Nicolson, 2004)	Not reported	Not reported	Not reported	Yes – drack, throld, greel, remshill	Not reported	No
Graded Nonword Reading Test (Snowling, Stothard, & McLean, 1996)	Not reported	Not reported	Report that 'all the nonwords can be pronounced according to regular letter-to-sound conversion rules'	Yes – sted, gromp, stansert, hinshink	Not reported	No
Illinois Test of Psycholinguistic Abilities, 3rd edition (Hammill, Mather, & Roberts, 2001)	Not reported	Not reported	Reports items are 'phonically regular'	Yes – stip, brump, flant, drooster	Not reported	No

(continued)

Appendix E (CONTINUED)
Review of Nonword Reading Tests: Nonwords

Test	Is GPC/nonword consistency taken into account?	Is rule frequency taken into account?	Are items orthographically legal?	Do the items contain parts that are real words?	Is each GPC tested in all orthographically legal word positions?	Do the items include pseudo-homophones?
MULTILIT Word Attack Skills Placement Test (MULTILIT, 2007)	Not reported	Not reported	States items are 'phonetically regular'	No	Not reported	No
Nonword Decoding Test (Turner, 2003)	Not reported	Not reported	Not reported	Yes – slimp, grash, blit	Not reported	Yes - sode
Nonword Reading Test (Martin & Pratt, 2001)	Not reported	Not reported	Reports that nonwords 'all conform to acceptable letter-sound correspondences'	Yes – hoff, floil, wrapple, bunset, theld, piggle	Not reported	No
Nonword Reading Test (Wren, 2003)	Not reported	Yes, for the irregular words from which the nonwords were created	Not reported	Yes – cread, biter	Not reported	No
Phonological Assessment Battery (Frederickson, Frith, & Reason, 1997)	Not reported	Not reported	Not reported	No	Not reported	No
Queensland University Inventory of Literacy (Dodd, Holm, Oerlemans, & McCormick, 1996)	Not reported	Not reported	Not reported	Yes – slet, framyip, stinter, dickshenree, acksident, domplimint	Not reported	Yes – bocks, enuf, acksident, ocksidgen

Appendix E (CONTINUED)
Review of Nonword Reading Tests: Nonwords

Test	Is GPC/nonword consistency taken into account?	Is rule frequency taken into account?	Are items orthographically legal?	Do the items contain parts that are real words?	Is each GPC tested in all orthographically legal word positions?	Do the items include pseudo-homophones?
Sutherland Phonological Awareness Test – Revised (Neilson, 2003)	Not reported	Not reported	Not reported	Yes – hipstan, mespuntal, splant, despintal	Not reported	No
Test of Word Reading Efficiency (Torgesen, Wagner, & Rashotte, 1999)	Not reported	Yes	Not reported	Yes – they advocate the inclusion of such items as showing growth in decoding skills	Not reported	Yes – knap, faw, pu, mest, ni
Wechsler Individual Achievement Test – 2nd Australian edition (The Psychological Corporation, 2002)	Not reported	Not reported	Reported that the nonwords 'mimic the phonological structure of the English language'	Yes – sluck, flid, waim, chag, thag, sorb, tellitry, pragment	Not reported	Yes – cind
Woodcock Johnson III (Woodcock, McGrew, & Mather, 2001)	Not reported	Not reported	Reports that 'letter combinations are phonologically consistent, or regular, patterns in English orthography'	Yes – blighten, splaunch	Not reported	No
Woodcock Reading Mastery Tests – Revised (Woodcock, 1998)	Not reported	Not reported	Not reported	Yes – din, glack, rayed	Not reported	Yes – rayed, knap