A review of the teaching of early reading in England commissioned by the UK Government recommended that synthetic phonics should be the preferred approach for young English learners. In response, all English schools have been told to put in place a discrete synthetic phonics programme as the key means for teaching high-quality phonic work. In this paper we analyse the evidence presented by the review to support the change to synthetic phonics. We show that the review provided no reliable empirical evidence that synthetic phonics offers the vast majority of beginners the best route to becoming skilled readers. We analyse the available empirical evidence in English, and show instead that the data support approaches based on systematic tuition in phonics. There is also evidence that contextualised systematic phonics instruction is effective. However, more research is needed, particularly with typically developing readers, in order to determine whether contextualised systematic phonics is more effective than discrete systematic phonics.

The importance of systematic phonics instruction in relation to the teaching of reading has been increasingly recognised by English-speaking countries. In 2000 the USA’s National Reading Panel advocated systematic phonics instruction as part of a balanced programme of reading teaching (National Institute of Child Health and Human Development [NICHD], 2000). In 2005, the Australian government enquiry concluded that systematic, direct and explicit phonics instruction was important as part of an integrated approach to reading teaching (Australian Government, Department of Education Science and Training, 2005). In England, the government enquiry into the teaching of early reading (hereafter the Rose Report) reached a narrower conclusion (Rose, 2006). Rose’s recommendation was that the systematic phonics approach selected by schools should be synthetic phonics. The Rose Report also put forward a ‘simple view of reading’ (Rose, 2006, Appendix by Stuart & Stainthorp). This adaptation of the classic ‘simple view’ put forward by Gough and Tunmer (1986) recommended clear differentiation between word recognition processes and language comprehension processes, so that teachers could assess performance separately in each and plan different kinds of teaching for each dimension. One author of this Appendix explains, ‘confounding the two dimensions
is not conducive to effective teaching in either’ (Stuart, 2006, p. 9). The simple model provided the basis for the recommendation in the Rose Report that systematic phonics should be taught ‘discretely’ (p. 70).

Synthetic phonics and the simple view of reading is being introduced immediately to all English schools. Hence, a universal intervention is being introduced on the basis of the evidence reviewed by Rose (2006). As noted by Rutter (2006), it is always desirable to determine the efficacy of an intervention under optimal research conditions such as a randomised controlled trial before launching an intervention universally. For programmes intended to make a real difference in the long term, such as synthetic phonics, the research evaluation should be long term, and it must be recognised that subgroups (for example, children with learning difficulties) may require something different. Our first question in this article is whether any of these desirable conditions were met by the evidence base considered by Rose (2006). Rose outlined his remit as ‘to examine … what best practice should be expected in the teaching of early reading and synthetic phonics’ (2006, p. 7). We therefore focus on the evidence that he gathered to meet that remit concerning best practice.

As will become evident, one difficulty in analysing the evidence base underpinning the report is that Rose did not consider empirical research evidence as the primary form of data relevant to recommending changes in national policy on the teaching of reading. His approach is formalised in paragraph 47 of the review:

The review’s remit requires a consideration of ‘synthetic’ phonics in particular … through examination of the available evidence and engagement with the teaching profession and education experts. Having followed those directions, and notwithstanding the uncertainty of research, there is much convincing evidence to show from the practice observed that … synthetic phonics … offers the vast majority of beginners the best route to becoming skilled readers … (p. 19, emphasis added)

Stuart (2006), a major adviser to Rose, also argued that the inconclusive nature of available research evidence required Rose to seek recourse to observation of selected teaching practices in order to make national recommendations: ‘the available research evidence is insufficient to allow reliable judgements of the relative effectiveness of implementing different approaches to systematic structured phonics teaching’ (Stuart, 2006, p. 11). Rose argued that empirical research evidence cannot provide the answers that he required, as ‘In recent years, there has been a convergence of opinion among psychologists investigating reading that little progress towards understanding how reading happens in the human mind is likely to be made’ (p.75, emphasis added). He stated, ‘focusing on the practice observed in the classroom and its supportive context, rather than debating the research, is [therefore] not without significance for this review’ (p. 62). Our second question in this article is therefore whether the empirical research base on the teaching of early reading is really so uncertain.

An alternative view would be that a great deal of progress has been made in understanding the acquisition and development of reading, and its neural underpinning in the human mind. To paraphrase Frith (1998), ‘reading literally changes
the brain’ (p. 1051). Importantly, it does so in somewhat different ways in different languages (Goswami, 2006). Understanding the particular demands made by a language like English is crucial for the successful early teaching of reading in English. This is because written English is different from many other alphabetic languages. The phonological complexity of syllable structures in English, coupled with the inconsistent spelling system, mean that direct instruction at levels other than the phoneme may be required in order to become an effective reader. In countries like Greece, Finland, Italy and Spain, syllable structure is simple (mainly consonant-vowel syllables) and there are 1:1 mappings between letters and sounds. Accordingly, most children are fluent readers long before the end of their first year of instruction (Seymour et al., 2003). For such languages, teaching reading by synthetic phonics can be extremely effective (Landerl, 2000). In English-speaking countries, such early fluency has proved elusive. Our third question in this article is what empirical studies can tell us about the effective teaching of reading in English. As will be illustrated, the complexity of reading acquisition in English makes it unlikely that the universal adoption of one method, synthetic phonics, without any evidence of proof of concept (for example, via randomised controlled trials), will deliver the results expected by Rose and by the UK Government.

The case for systematic phonics

One of the most significant contributions to questions about research evidence and the teaching of reading was the American National Reading Panel (NRP) report on teaching children to read in English (NICHD, 2000). This was an extensive meta-analysis of research evidence which addressed a number of questions about early literacy, including: ‘Does systematic phonics instruction help children learn to read more effectively than non-systematic phonics instruction or instruction teaching no phonics?’ (ch. 2, p. 92), and ‘Are some types of phonics instruction more effective than others? Are some specific phonics programs more effective than others?’ (ch. 2, p. 93). As far as differences between analytic and synthetic phonics programmes were concerned, the NRP concluded that ‘specific systematic phonics programs are all significantly more effective than non-phonics programs; however, they do not appear to differ significantly from each other in their effectiveness although more evidence is needed to verify the reliability of effect sizes for each program’ (NICHD, 2000, ch. 2, p. 93). In the NRP, the effect sizes for the phonics approaches classified as ‘synthetic’ versus ‘analytic’ were statistically equivalent.

More recently, England’s Department for Education and Skills (DfES) commissioned a large-scale study of approaches to the teaching of reading. The methodology of the NRP was refined to produce a meta-analysis focused on studies that were randomised controlled trials (RCTs). On the basis of their study Torgerson et al. (2006, p. 49) concluded that ‘There is currently no strong RCT evidence that any one form of systematic phonics is more effective than any other’. As RCTs are one form of the ‘optimal research conditions’ noted by Rutter (2006), this is an important finding.
Two rigorous forms of reviewing large numbers of research studies, the meta-analyses conducted by the NRP and the meta-analysis restricted to RCTs by Torgerson et al. (2006), converge in concluding that the systematic teaching of phonics is crucial for literacy acquisition. This conclusion was also recognised by Rose (2006), who stated, ‘Having considered a wide range of evidence, the review has concluded that the case for systematic phonic work is overwhelming …’ (p. 20). However, he continued, ‘… is overwhelming, and much strengthened by a synthetic approach’. He then listed the core components of any systematic phonics programme, namely learning grapheme-phoneme correspondences, and learning to blend and segment. Despite this common ground between systematic phonics and synthetic phonics, it is stated that ‘synthetic phonics … offers the vast majority of beginners the best route to becoming skilled readers’ (p. 19). Rose wrote, ‘analytic phonics is good, but synthetic phonics is better’ (p. 19, quote attributed to Rhona Johnston). In the section of the Rose Report outlining what best practice should comprise (pp. 15–28), the footnotes refer to various forms of empirical evidence such as the NRP and Torgerson et al. reviews, and one published empirical study on synthetic phonics, Johnston and Watson (2005). The section of the Rose Report presenting evidence from practice (pp. 61–69) focuses on visits made by Rose and his team to the Scottish educational authority of Clackmannanshire.

The case for synthetic phonics

The studies known in the UK as the ‘Clackmannanshire studies’ have been written up in a number of forms by Johnston and Watson (Johnston & Watson, 2003, 2004, 2005; Watson & Johnston, 1998). Subsequent to the adoption of a reading tuition programme called by these authors ‘synthetic phonics’, children were reported to make considerable gains in reading. The Rose Report summarised these gains as follows:

At the end of Primary 7 [Year 6 in England, i.e. age 10 to 11], word reading was 3 years 6 months ahead of chronological age, spelling was 1 year 8 months ahead and reading comprehension was 3.5 months ahead. However, as mean receptive vocabulary knowledge ... was 93 at the start of this study ... this may be an underestimate of the gains with this method. (p. 61)

Johnston and Watson’s (2005) work is not a research paper, rather it is a summary of one of their studies written for a government website. However, one of their publications concerning their Clackmannanshire data sets has been peer-reviewed (Johnston & Watson, 2004), and Experiment 2 in that paper are the same data discussed in Johnston and Watson (2005), the main evidence used in the Rose Report. We therefore focus on analysing the data presented in Experiment 2 of Johnston and Watson’s (2004) paper.

Johnston and Watson (2004) reported two studies, although chronologically Experiment 2 was conducted before Experiment 1. Experiment 1 will not be analysed in depth here (see Goswami, 2008, for a thorough analysis), as it was not mentioned as evidence in the Rose Report. Experiment 2 was an intervention study,
with both types of intervention administered by the authors themselves. Intervention was extra to normal classroom tuition, and began six weeks after school entry. This extra training was given twice weekly for 10 weeks, and was based on a set of 114 printed words. A control group (‘no-letter training group’) was ‘exposed to the print vocabulary shown to the other two groups, but a “look-and-say whole word” approach was adopted’ (p. 346). One intervention group (‘accelerated letter learning group’, also referred to as the ‘analytic phonics’ group) was taught two letters per week in the initial position in the 114 words. ‘Various games were played where the pupils matched pictures and words, however their attention was drawn only to the initial sounds of the words and the letters representing those sounds’ (p. 347).

The second intervention group (‘synthetic phonics group’) was also taught two letters per week, but learned these letters in all positions of the words (‘accelerated learning and blending of the letter sounds in initial, middle and final positions of words’, p. 347). The experimenter demonstrated how the words could be read by saying each letter sound in a word distinctly from left to right, joining them together smoothly without pausing between each sound. The children practised reading each word in this way, first of all with the experimenter, and then independently … using lower case letters and a magnetic board. (p. 347)

Therefore, only the synthetic phonics group was taught to read the 114 target words independently.

Johnston and Watson (2004) provide post-test data by group for the participants three months after the end of the intervention, and nine months after the end of the intervention. Three months after the intervention ended, the synthetic phonics group had a mean reading level of 5 years 7 months (mean age in all three groups at this point was 5 years 5 months). The accelerated letter learning group and the no-letter control group each had a mean reading age of 5 years 3 months, a significantly lower level of attainment. The superior performance of the synthetic phonics group was maintained nine months later, by which time the other two groups had also been taught (by their classroom teachers) to analyse letters in other positions in words—yet had failed to ‘catch up’ (p. 351). The authors concluded:

Experiment 2 … [controlled] for speed of letter learning; the synthetic phonics group still read and spelt better than the analytic phonics group. It is concluded that synthetic phonics was a more effective approach to teaching reading, spelling and phonemic awareness than analytic phonics. (p. 351)

Reading comprehension was reported to be equivalent for all three groups (p. 251). However, the study design is not rigorous enough to enable a conclusion about the superiority of one teaching method against the other. This is because the absolute amount taught to the children in each intervention group is confounded with the labels given to the teaching methods. Johnston and Watson (2004) define ‘analytic phonics’ as ‘children [learning] letter sounds in the context of words that they have been taught to recognize by sight’ (p. 329). A stringent comparison of analytic versus synthetic teaching methods would require each group to be taught the same total number of letter-sounds in the context of all positions in words, and to be taught the same component skills of sounding out, blending and decoding the 114 target
words. To compare teaching methods, the required design should contrast whether this teaching occurred in the context of sight words (i.e. analytically) or began from discrete sounds (i.e. synthetically). In Johnston and Watson’s study, only the synthetic phonics group were taught sounding out, blending and reading of the 114 words in the target list. This is not an optimal research design for exploring the research question of whether synthetic phonics methods are superior to analytic methods (Experiment 1, which is not cited by Rose, did not use a rigorous research design either; see Goswami, 2008). It is currently unknown whether any differences would be found between children’s progress if the correct research design were to be used. Regrettably, the Rose Report did not recognise that such a study was required before recommendations about national policy could be made, nor indeed that an RCT of the method was required before synthetic phonics was rolled out nationally. This is particularly important in the light of the UK experience with Sure Start (part of the UK equivalent of No Child Left Behind; see Rutter, 2006). The efficacy of Sure Start was not determined under optimal research conditions before its launch, and early evaluations did not find evidence of clear benefit.

It is important to note that the Rose Report did accept that the methodology of the Clackmannanshire study ‘received some criticism by researchers’ (p. 61). Rather than presenting these criticisms, the report discussed a visit by Rose and his team to Clackmannanshire:

The visit provided the review with first-hand evidence of very effective teaching and learning of phonic knowledge … focusing on the practice observed in the classroom and its supportive context, rather than debating the research, is therefore not without significance for this review. (p. 62, emphasis added)

Rose tells us, ‘One teacher said “I have never seen results like this in 30 years of teaching”’ (p. 63). Such anecdotes share similarities with those supplied by teachers who have received training in various commercial phonics packages ‘THRASS² is like a new religion! I have seen the light—the answer to how to teach English spelling’ (THRASS, 1999); ‘Jolly Phonics has given our children the chance to succeed in reading and writing’ (Jolly Phonics Case Study, 1999). Regarding such packages, the Rose Review stated, ‘all these programmes were highly systematic and the perceived, sharp differences that divided their advocates appeared to make little difference to the claimed success rates’ (p. 20). Curiously, it is not recognised that exactly the same logic applies to the Clackmannanshire studies.

The case against synthetic phonics

Recent cross-language research suggests that there are two critical factors that affect the acquisition of efficient phonics skills across languages. One is the phonological complexity of the language. Most of the world’s languages have words made up of syllables with a simple or consonant–vowel (CV) structure. English words made up of simple CV syllables include ‘yoyo’ and ‘baby’. Languages with a simple CV syllable structure include Italian, Spanish and Finnish. Words in these languages tend to be longer than words in English, but easier to segment into phonemes. This
is because the onset-rime level of analysis of the spoken syllable and the phoneme level of analysis of the spoken syllable are identical. This is shown in Figure 1. Children seem to acquire phonic recoding skills much faster when the phonology of their language has a simple syllabic structure. In English, only 5% of monosyllables are CV (De Cara & Goswami, 2002). The primary syllable type in English is CVC (43% of monosyllables), followed by CVCC (21%) and CCVC (15%). Note, however, that languages like German do not have a simple syllabic structure either. In fact, words in German are frequently exactly the same as words in English. Words like sand (Sand), ball (Ball), wine (Wein) and mouse (Maus) are phonologically identical. For words like these, onset-rime segmentation is not equivalent to phonemic segmentation.

Nevertheless, German children learn phonics much faster than English children (Frith *et al.*, 1998; Landerl, 2000). This is because of the second critical factor.
affecting the efficient acquisition of phonics skills, the consistency of the symbol-to-sound mapping (Ziegler et al., 1997). In English, one letter or letter cluster can have multiple pronunciations (e.g. ‘cough’, ‘rough’, ‘though’; this is also true of languages like Danish). In other languages, letters are always pronounced in the same way (for example, German, Greek, Italian, Spanish). In English, a single speech sound (phoneme) can have multiple spellings (e.g. the phoneme /ɔ/ in hurt, dirt, Bert). In other languages, a phoneme may be almost always spelled in the same way (e.g. Italian, Serbo-Croatian). English is exceptionally inconsistent, because it suffers from a large amount of inconsistency in both reading and spelling. This inconsistency inhibits the rapid acquisition of grapheme–phoneme recoding skills (Wimmer & Goswami, 1994; Goswami et al., 1998, 2001; Seymour et al., 2003). The acquisition of phonic recoding skills during the first year of reading tuition across the European Community nations is shown in Table 1, adapted from Seymour et al. (2003). It should be noted that the English-speaking children tested were schooled in Scotland using predominantly phonics methods. The table shows clearly that children learning to read highly transparent orthographies where the spoken language has a simple CV syllable structure are extremely efficient in grapheme–phoneme decoding within months of entering school.

Direct cross-language comparisons of teaching beginning readers by synthetic phonics support the empirical evidence shown in Table 1. Landerl (2000) compared English children taught via synthetic phonics to German children taught using similar methods. She commented that in German initial reading was typically taught via ‘a straightforward phonics teaching regime … through heavy emphasis on teaching of letter–sound correspondences and blending and sounding out words’ (p. 240). She therefore sought out schools in England following a similar regime, in order to compare the progress made by English children receiving ‘straightforward

<table>
<thead>
<tr>
<th>Language</th>
<th>Familiar real words</th>
<th>Pseudo-words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greek</td>
<td>98</td>
<td>92</td>
</tr>
<tr>
<td>Finnish</td>
<td>98</td>
<td>95</td>
</tr>
<tr>
<td>German</td>
<td>98</td>
<td>94</td>
</tr>
<tr>
<td>Austrian German</td>
<td>97</td>
<td>92</td>
</tr>
<tr>
<td>Italian</td>
<td>95</td>
<td>89</td>
</tr>
<tr>
<td>Spanish</td>
<td>95</td>
<td>89</td>
</tr>
<tr>
<td>Swedish</td>
<td>95</td>
<td>88</td>
</tr>
<tr>
<td>Dutch</td>
<td>95</td>
<td>82</td>
</tr>
<tr>
<td>Icelandic</td>
<td>94</td>
<td>86</td>
</tr>
<tr>
<td>Norwegian</td>
<td>92</td>
<td>91</td>
</tr>
<tr>
<td>French</td>
<td>79</td>
<td>85</td>
</tr>
<tr>
<td>Portuguese</td>
<td>73</td>
<td>77</td>
</tr>
<tr>
<td>Danish</td>
<td>71</td>
<td>54</td>
</tr>
<tr>
<td>Scottish English</td>
<td>34</td>
<td>29</td>
</tr>
</tbody>
</table>
phonics’ with that of German children. The phonics method chosen was a commercial programme in use in the UK, Jolly Phonics (Lloyd, 1998). An additional contrast was with the ‘standard’ method of reading tuition in England at that time (preceding the National Literacy Strategy of 1998), which was classified as ‘a mixed approach providing a combination of whole word and phonics methods’ (p. 241)—‘children in this school received a phonics lesson once a week’ (p. 247). Thus two English schools adopting contrasting methods to the teaching of initial reading were compared to a German school utilising a synthetic phonics approach. The outcome measure was non-word reading at grades 1, 2, 3 and 4 in each case.

The important study was Experiment 2, for which Landerl herself carried out the testing of all participating children (in Experiment 1 the synthetic phonics group was tested by advocates for the Jolly Phonics method, and the data from these children were compared to children from a different education authority tested years previously by Goswami’s research group). In Experiment 2 there were 111 children in the ‘English standard’ group, 87 children in the ‘English phonics’ group, and 102 children in the German group. Non-word reading accuracy at Grades 1 (age 6) and 2 (age 7) was found to be equivalent for the two sets of English children. Both groups were significantly poorer than the German children. At Grade 3, non-word reading accuracy was much better, with the English children now making relatively few errors (28% for the standard group, and 7% for the phonics group). This was the first time that the two groups differed in non-word reading. By Grade 4, the difference had disappeared. Landerl concluded:

the main difference between the two English groups was in Grade 3. In Grade 1, both groups of English children committed a high number of errors, and in Grades 2 and 4 the difference was comparably small, because the English standard children performed almost as well as the English phonics children. (p. 250)

The equivalence of the English standard and English phonics children is remarkable given that the English standard teachers reported only administering one phonics lesson a week.

One objection to cross-language comparisons is that cultural differences may explain differences that are attributed to language per se. One empirical solution is to look for cultures where bilingualism is common. In Wales, for example, children grow up with either Welsh or English as their native tongue. Although they live in the same geographical areas, their parents can choose whether to have them schooled in Welsh or English. The schools themselves follow a similar (phonics-based, e.g. Jolly Phonics) reading curriculum, are in the same catchment areas (in some towns, they are side by side), and are administered by the same education authorities. Welsh is an extremely consistent orthography. The mappings from graphemes to phonemes are unambiguous. However, the syllable structure is complex. Welsh syllables can have complex onsets, as in streic and cnau, or complex codas (consonant phonemes after the vowel), as in sinc and peint. Hence, in terms of orthographic transparency, Welsh and English differ markedly, whereas in terms of phonological complexity, they are extremely similar. Spencer and Hanley (2003) utilised these unique features of primary schools in Wales in order to carry out cross-language comparisons of
reading acquisition when cultural factors were held constant. The children were in their second year of formal reading instruction, and were on average six years old.

Spencer and Hanley reported that the Welsh-speaking children were significantly more accurate in both word reading and non-word reading compared to the English-speaking children. For real words, the Welsh children read on average 81% of words correctly, compared to 59% correct for the English children. For non-words, the Welsh children read on average 78% of words correctly, compared to 44% correct for the English children. This was not due to a disparity in the difficulty of the items used in English and Welsh. For the non-words, for example, the Welsh-speaking children decoded 68% of the English items correctly. The English-speaking children only decoded 33% of the Welsh items correctly. The Welsh children had also developed superior phonemic awareness (measured by a phoneme counting task). Recall that the phonological complexity of words in English and Welsh is similar. The finding that phonemic awareness was better in Welsh hence reflects the importance of learning 1:1 grapheme–phoneme correspondences for the rapid development of phonemic awareness. Spencer and Hanley concluded that their findings ‘provided strong evidence that reading acquisition is heavily influenced by the transparency of the alphabetic writing system’ (p. 14). When followed up a year later, the Welsh-speaking children were still significantly better at word reading, non-word reading and phoneme counting than the English-speaking children. Children learning to read a transparent orthography will hence acquire reading skills more quickly, even when teaching methods are based on synthetic phonics. When tested again three years later, the differences between the Welsh-speaking and English-speaking children had disappeared (Hanley et al., 2004). This was the sixth year of formal reading instruction, and the children were now 10 years old. Hanley et al. noted, however, that at this test point the English children had significantly better reading comprehension skills than the Welsh children.

Clearly, cross-language comparisons suggest that linguistic factors rather than teaching method per se explain the slower acquisition rates shown by English-speaking children. However, one point made by Johnston and Watson (2005) was that synthetic phonics should be taught first, before any other kind of reading tuition was received (‘there is evidence that synthetic phonics is best taught at the beginning of Primary 1’ [the first year of formal schooling], p. 9). A comparable training study in this respect focusing on English only was reported by Walton et al. (2001). They contrasted two types of phonics training, one based on phonemes (called letter recoding) and the second based on larger units (called rime analogy). These two training methods are characteristic of the wider literature comparing ‘synthetic’ with ‘analytic’ methods. This literature usually regards analytic phonics as going from ‘whole to part’, focusing initially on larger grain sizes (Moustafa & Maldonado-Colon, 1998). Synthetic phonics is regarded as going from ‘part to whole’, focusing initially on the smallest grain size of the phoneme. For the children studied by Walton et al., who were pre-readers, this was the first direct tuition in reading received by the participating children. The two groups were taught the same pre-reading skills (phonological awareness of initial, medial and final phonemes,
rhyming and the letter–sound correspondences), and the treatments were controlled so that the only differences between the training concerned whether the first steps of reading were taught using a letter–sound recoding strategy or a rime analogy strategy. Two training sessions of 25 minutes were given twice a week for 11 weeks.

Reading outcomes were then assessed. For word reading, a significant advantage was found for the rime analogy group. For non-word reading, the two groups performed at an equivalent level. In each case, the trained children outperformed a control group who had received normal classroom teaching. A set of longitudinal post-tests given four months later showed equivalent patterns. Walton et al. concluded that ‘the study demonstrated that Grade 1 pre-readers with weak letter–sound and phonological skills will develop reading ability relatively quickly if given experience with the rime analogy or letter recoding strategies and the related pre-reading skills’ (p. 178, emphasis added). The essential finding was that as long as tuition was systematic, either large-unit (‘analytic’) or small unit (‘synthetic’) phonics lead to similar gains.

Although carried out with very different samples, the experimental studies by Walton et al. (2001), Landerl (2000) and Spencer and Hanley (2003) all reach the same conclusion. The conclusion is that no one method of teaching phonics to children learning to read in English appears to be superior to any other method. This is exactly the same conclusion as was reached by the meta-analyses reported by the NRP and by Torgerson et al. (2006). There is no empirical research base to justify the Rose Report’s recommendation that the teaching of reading in England must rely on synthetic phonics.

How do we teach early reading in English?

What, then, should teachers of early reading do? An empirical approach to this question is to analyse the teaching methods featured in the phonics instruction studies that were included in the NRP and Torgerson et al. (2006) reviews. Table 2 categorises the 43 studies included in the two reviews by the pedagogical emphasis of the most effective instruction. Table 2 shows that the largest number of studies featured instruction where phonics was contextualised at sentence and text level as part of the teaching programmes. Interventions with an emphasis on discrete phonics, where comprehension teaching and work with whole texts were covered separately from the phonics instruction, were smaller in number but were also shown to be effective. Whole language, whole-word teaching and broader approaches were also shown to be effective in some studies.

The studies categorised in Table 2 show that a range of different teaching approaches, including contextualised systematic phonics instruction and discrete systematic phonics instruction, can be effective. However, in view of the recommendation by the Rose Report that phonics teaching should be discrete, a relevant question is, which teaching approach is more effective, contextualised phonics instruction or discrete phonics instruction? In order to answer this question there is a need to identify studies which compare directly the two kinds of
The selection criteria we used for studies which would address this question were as follows:

1. explicit comparison of contextualised systematic phonics instruction versus discrete systematic phonics instruction;
2. rigorous experimental design, preferably randomised controlled trial;
3. sample of typically developing readers;

Table 2. Methods of teaching reading implemented in research studies

<table>
<thead>
<tr>
<th>Pedagogical emphasis of effective interventions</th>
<th>Studies</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrete phonics work</td>
<td>Bond et al. (1995–96) (included spelling work, songs and books); Foorman et al. (1991); Haskell et al. (1992) (onset-rime group outperformed others); Johnston and Watson (2004); Lovett et al. (2000) (sub-syllabic, particularly focused on rime); Mantzikopoulos et al. (1992); Oakland et al. (1998); Snider (1990); Stuart (1999)</td>
<td>9</td>
</tr>
<tr>
<td>Whole language teaching, including literature teaching</td>
<td>Eldredge (1991) (included some systematic phonics); Freppon (1991); Griffith et al. (1992); Klesius et al. (1991); Wilson and Norman (1998)</td>
<td>5</td>
</tr>
<tr>
<td>Contextualised phonics instruction: Phonics instruction contextualised with sentence-level and/or text-level work such as reading of connected text. Reading recovery and modified reading recovery type programmes included.</td>
<td>Berninger et al. (2003); Blachman et al. (1999); Brown and Felton (1990); Evans and Carr (1985); Foorman et al. (1997); Foorman et al. (1998); Greaney et al. (1997); Martinussen and Kirby (1998); Santa and Hoen (1999); Tunmer and Hoover (1993); Umbach et al. (1989); Vickery et al. (1987)</td>
<td>12</td>
</tr>
<tr>
<td>Whole word teaching</td>
<td>Gittelman and Feingold (1983) (included intersensory phonics method); Lovett et al. (1990); Lovett and Steinbach (1997) (included metacognitive training for strategy monitoring); O’Connor and Padelia (2000)</td>
<td>4</td>
</tr>
<tr>
<td>Broad approach and/or curriculum focus</td>
<td>Gersten et al. (1988); Lum and Morton (1984) (focus on spelling more than reading); Marston et al. (1995); Vandervelden and Siegel (1997) (argued against sounds pronounced in isolation)</td>
<td>4</td>
</tr>
<tr>
<td>No statistically significant gains overall for any approach</td>
<td>Lovett et al. (1989) (showed different gains for decoding skills and OWLS approaches); Silberberg et al. (1973) (included synthetic phonics); Skailand (1971); Torgesen et al. (1999) (advocates combined word-level and comprehension teaching); Torgesen et al. (2001) (Embedded Phonics as effective as Auditory Discrimination in Depth); Traweek and Berninger (1997)</td>
<td>6</td>
</tr>
<tr>
<td>Limited information about teaching approaches</td>
<td>Fulwiler and Groff (1980); Leach and Siddall (1990) (work with parents); Leinhardt and Engel (1981)</td>
<td>3</td>
</tr>
</tbody>
</table>
4. Standardised measures of comprehension in addition to standardised measures of word reading included.

None of the studies met all four criteria. This is because no studies of this kind have been carried out with typically developing readers. Only two studies met three of the criteria: the studies of children showing delayed development by Berninger et al. (2003) and Torgesen et al. (2001).

With regard to the methodological criteria for selection of rigorous experimental trials, we adopted Torgerson et al.’s (2006) quality assessment criteria, used to identify the RCTs included in their main analysis (and in two cluster trials), which were as follows: (a) reporting of method of random allocation; (b) sample size justification; (c) intention to teach analysis, where all participants are analysed in their original randomised groups as opposed to excluding participants after randomisation; (d) blinded assessment of outcome.

Of the two studies that met our criteria for comparing discrete phonics instruction with contextualised phonics instruction Torgesen et al. (2001) was the only study that also met one of the experimental trial criteria, the intention to teach analysis. Sixty children between the ages of eight and ten who had been previously identified as learning disabled were recruited for participation in the study by Torgesen et al. (2001). The children were randomly assigned to one of two groups: (a) auditory discrimination in depth (ADD); (b) embedded phonics (EP). The teaching was provided by educational therapists on a one-to-one basis in two 50-minute sessions each day of the week. Teaching was provided over a period of 8–9 weeks until 67.5 hours of instruction were accomplished.

The ADD programme had three major goals. The first goal was to provide a basis for accurate discriminations between phonemes. The second goal was to teach children to use their knowledge of phonemes to monitor and represent sequences of sounds in spoken syllables. The third goal was to teach self-monitoring skills that allowed children to discover methods to correct themselves during their phonics decoding attempts. All 44 English consonant and vowel phonemes were taught and the programme provided extensive practice in reading and spelling individual words.

The EP programme was designed to provide direct, explicit instruction in word-level reading skills while providing extensive opportunities to read and write meaningful text. The programme included practice in reading sight words; practice in spelling sight words; word games for fluency with sight words; phonics ‘mini lessons’; oral reading of trade books or basal readers; and writing activities using sight words.

Outcomes were very similar for both instructional methods. The effects of intervention were both substantial and stable over the course of the two-year follow-up period. Torgesen et al. (2001) argued that the goal of any intensive intervention is to produce large changes in reading ability that were maintained over time and that ‘the clear conclusion from this research is that the EP and ADD conditions were equally successful’ (p. 51). In view of the fact that two approaches that varied substantially were equally effective, Torgerson et al. argued that within structured language approaches the components of instruction could be arranged according to teacher and student preferences.
Berninger et al. (2003) did not meet any of the Torgerson et al. (2006) quality assessment criteria but will be assessed here because it was a cluster RCT. The study included the hypothesis that ‘instruction that is aimed at all the necessary components of a functional reading system is more effective than instruction that is aimed at an isolated component of the reading system’ (Berninger et al., 2003, p. 103). In the study, second-grade teachers in eight schools serving diverse student populations were asked to refer their poorest readers. A test battery covering vocabulary, word identification and word attack was then administered by project personnel in order to determine the final sample. Forty-eight pairs of children were randomly assigned to four conditions: (a) explicit and reflective word recognition; (b) explicit and reflective reading comprehension; (c) combined explicit word recognition and explicit reading comprehension; or (d) treated control that only practised reading skills without any instruction. The combined condition featured explicit instruction in the alphabetic principle and explicit language cueing at the word, sentence and text level to stimulate the text-based component of reading comprehension. The teaching consisted of twice-weekly sessions of 20 minutes each for a total of 24 lessons. The word recognition training began with 10 minutes of explicit instruction in the alphabetic principle, both in and out of word context. In any one session, 12 to 24 correspondences were introduced or reviewed, and altogether, 107 correspondences were practised four or five times during the course of the tutorial. During the next six minutes of training, children practised applying these correspondences to decoding, which involved pronouncing single words. During the last 10 minutes teachers provided explicit language cueing at the word, sentence and text levels (Berninger, 1998), in order to stimulate the text-based component of reading comprehension. Levels of language cueing to develop text-based comprehension always contained word-level prompts, sentence-level prompts and text-level prompts. An example of a text-level prompt included: ‘Tell the plot or main events in the story so far’ (p. 107).

Berninger et al. reported that (c), the combined word recognition and reading comprehension treatment, was the most effective in increasing phonological decoding as shown by gains in non-word and real word reading. Given that most at-risk readers have difficulty with phonological decoding, this is an important finding. However, all three explicit treatments and the practice control treatment led to significant improvements in reading comprehension. Berninger et al. (2003) theorised that explicit instruction in comprehension combined with explicit word recognition training facilitated learning to decode written words because it helped to develop broad-based metalinguistic awareness. This was thought to generalise across levels of language in the functional reading system.

The main conclusion to be drawn from this analysis of phonics instruction studies is that a range of ways of teaching systematic phonics is effective. Considering contextualised systematic phonics instruction versus discrete systematic phonics instruction, there is a need for more research to compare the effectiveness of the two approaches. In particular, no studies of typically-developing readers currently exist. The two studies that have most rigorously investigated this question differ in their
findings. Torgesen et al. (2001) found that contextualised instruction and discrete instruction were equally effective; Berninger et al. (2003) found that contextualised phonics instruction was more effective for word decoding but that there was no difference for reading comprehension.

Our analysis of empirical evidence does not support the advice based on the Rose Report’s ‘simple view of reading’ that schools should put in place a ‘discrete programme as the key means for teaching high-quality phonic work’ (DfES & PNS, 2006). Rather, the studies reviewed here suggest that many forms of systematic phonics instruction are successful. Contextualised methods in which systematic phonics are carefully integrated with whole text work and comprehension activities are effective (Berninger, 2003, method c). According to our analysis, the key features of phonics (learning grapheme–phoneme correspondences, learning to segment and blend) as identified by Rose can be taught by a range of systematic methods.

Conclusion

As a result of the recommendations of the Rose Report (Rose, 2006), the UK Government’s Primary National Strategy (PNS) Literacy Framework and its guidance, which affects all early years and primary teachers in England, was changed. The guidance which accompanies the PNS literacy framework makes it clear that schools and early years settings should ‘put in place a systematic, discrete programme as the key means for teaching high-quality phonic work’ and that ‘Shared and guided reading sessions should not be used to replace discrete phonics teaching but they can provide opportunities to reinforce children’s developing phonic knowledge and skills, in the context of achieving the ultimate goal of the sessions, which is the development of comprehension’ (DfES & PNS, 2006, p. 7). The UK Government in England, under the auspices of the Primary National Strategy, has also published a detailed programme called ‘Letters and Sounds: principles and practice of high quality phonics’ (Primary National Strategy, 2007) which teachers are expected to follow. Following the teaching of general orientation to sound discrimination in the nursery years, daily lessons for a six-week period feature ‘discrete phonics’ teaching. Teachers must ‘teach at least 19 letters, and move children on from oral blending and segmentation to blending and segmenting with letters’ (p. 48). Application of this knowledge during the Letters and Sounds lessons is limited to ‘Read or write a caption (with the teacher) using one or more high-frequency words and words containing the new letter (week 3 onwards)’ (p. 49). This is followed by further discrete teaching, lasting for up to 12 weeks. The purpose of this phase ‘is to teach another 25 graphemes, most of them comprising two letters (e.g. \(\text{oa}\)), so the children can represent each of about 42 phonemes by a grapheme’ (p. 74). Application at this stage is to ‘Read or write a caption or sentence using one or more tricky words and words containing the graphemes’ (p. 75). This pattern of a limited context for application of grapheme–phoneme correspondences continues through year one (age 5 to 6) until year two (age 6 to 7), at which point phonics instruction moves to an emphasis on spelling.
As demonstrated here, the Rose Report’s conclusion that synthetic phonics should be adopted nationally in England as the preferred method for the teaching of reading is not supported by empirical research evidence. Rather, as reviewed here, the available research evidence supports the importance of systematic tuition in phonics at a variety of grain sizes (e.g. phoneme, onset-rime). There is also evidence that systematic phonics instruction that is contextualised can be effective although there is a need for more research. As Torgerson et al. (2006) argued, the current empirical database suggests that there is no evidence that ‘any one form of systematic phonics is more effective than any other’ (p. 49).

The main research study which seems to have influenced the conclusions drawn by Rose (2006), namely Johnston and Watson (2004), does not enable any conclusions to be made concerning the optimal method of phonics instruction. In particular, it does not support the claim that analytic methods (from word to sound) are inferior to synthetic methods (from sound to word). This is because the children were not taught the same material or the same component skills by the two methods. Importantly, the findings reported by Johnston and Watson (2004) do not bear at all on whether phonics should be taught discretely or in isolation.

The rationale for the change to synthetic phonics in England was also, in part, built on a critique of the previous National Literacy Strategy ‘searchlights’ model. This model advocated the use of four ‘strategies’ as a means to understanding text: ‘phonic (sound and spelling); knowledge of context; grammatical knowledge; word recognition and graphic knowledge’ (Department for Education and Employment [DfEE], 1998, p. 4). In the Appendix to the Rose Report, Stuart and Stainthorpe (Rose, 2006, Appendix) argue, on the basis of their analysis of cognitive psychological research, in favour of a ‘simple view of reading’. One of the most significant differences between their simple view and the ‘searchlights’ model is that word recognition processes and language comprehension processes are seen as two separable dimensions of reading for the purposes of teaching: ‘The case for change that we discuss … rests on the value of explicitly distinguishing between word recognition processes and language comprehension processes’ (Rose, 2006, p. 74). However, reading is not simple. Reading is one of the most complex achievements of the human brain. Human brains that learn to read English may in fact develop extra neural architecture that is not developed by brains learning to read more consistent alphabetic orthographies (Goswami & Ziegler, 2006). In written language, as in spoken language, the ultimate aim is communication and comprehension. We argue that teachers are more likely to help children to achieve this aim if government recommendations for practice are built on a rigorous synthesis of the full range of evidence, including research about different languages and effective reading teaching.

Notes
1. Significant parts of the research evidence referred to in this article, including its synthesis in an abbreviated version of the current section entitled ‘The case for synthetic phonics’, were submitted to the Rose Review as written evidence.
THRASS is the acronym for Teaching Handwriting, Reading and Spelling Skills (THRASS UK Ltd).

References


