British Journal of Psychology (2009), in press
© 2009 The British Psychological Society

The British Psychological Society

www.bpsjournals.co.uk

Predicting children's word-reading accuracy for common English words: The effect of word transparency and complexity

Ken Spencer*
Hull University, Hull, UK

The effects of printed word frequency and transparency measures on single word reading accuracy were examined in 105 six-year-old children. The results indicated that it may be necessary to re-appraise notions of orthography-to-phonology correspondences for children of this age. The influence of orthographic neighbourhood size appeared to derive from word frequency and graphemic complexity. The results also indicated that sonograph frequency was more predictive of reading accuracy than the GPC rules and weighted correspondences currently embodied in dual route and connectionist models of skilled reading.

Languages can be arranged along a continuum, for spelling and reading, from transparent (one-to-one mappings of letters and sounds) to opaque (one-to-many mappings). English and French are considered to be opaque languages for both; Greek and German tend to be more transparent for reading rather than spelling; and Italian and Finnish are considered to be transparent for both (Goswami, Porpodas, & Wheelwright, 1997). This variation between orthographies leads to processing differences for naming and lexical decision, and is reflected in the ease with which foundation literacy skills are acquired in transparent orthographies that have consistent correspondences (Cossu, Shankweiler, Liberman, Katz, & Tola, 1988; Frost, Katz, & Bentin, 1987; Katz & Frost, 1992; Oney & Goldman, 1984).

For normal and dyslexic adults, transparent languages offer a processing advantage (Paulesu *et al.*, 2000, 2001), and for normal and dyslexic children, accuracy levels in opaque languages are lower, and reading speed slower (Cossu, Gugliotta, & Marshall, 1995; Ellis *et al.*, 2004; Frith, Wimmer, & Landerl, 1998; Landerl, Wimmer, & Frith, 1997; Spencer & Hanley, 2003). Seymour, Aro, and Erskine (2003) studied 13 European orthographies, and found that for young children, when a threshold of low orthographic transparency is exceeded, there is an abrupt negative effect on the rate of literacy acquisition.

Correspondence should be addressed to Dr Ken Spencer, Centre for Educational Studies, Hull University, Hull HU6 7RX, UK (e-mail: k.a.spencer@hull.ac.uk).

DOI:10.1348/000712609X470752

Although naming, lexical decision, and literacy acquisition studies have compared languages that vary significantly in their orthographic transparency, this variation has not always been quantified in objective terms. For example, Seymour *et al.* (2003) used a hypothetical classification of European orthographies derived from estimates of the variability by team members. However, more detailed and objective approaches have been used to measure the word body transparency of French (Ziegler, Jacobs, & Stone, 1996) and English (Ziegler, Stone, & Jacobs, 1997), for both reading and spelling. In these studies, French was found to be 88% consistent for reading, whereas English was 69% consistent; but for spelling, French was 21% consistent and English 28%.

Transparency has also been studied at the grapheme–phoneme level for English reading (Berndt, Reggia, & Mitchum, 1987; Gontijo, Gontijo, & Shillcock, 2003; Spencer, 2009) and spelling (Hanna, Hanna, & Hodges, 1966; Spencer, 2009). Gontijo *et al.* proposed that at this level the orthographic transparency of a language could be calculated by dividing the total number of grapheme–phoneme correspondences (GPCs) in a language by the total number of graphemes for reading and by the number of phonemes for spelling.

Within opaque languages, the transparency of individual words varies and influences literacy acquisition and processing. Venezky (1970) introduced a series of rules for the pronunciation of English orthography, based on his grapheme–phoneme analysis of 20,000 words, and it was found that transparent regular words, as defined by his rules, were read aloud by adults with shorter naming latencies than opaque exception words (Baron & Strawson, 1976). There have since been numerous studies demonstrating longer naming latencies for exception words (e.g. Hino & Lupker, 2000; Stanovich & Bauer, 1978; Waters & Seidenberg, 1985).

Developmental and computer models of English reading and spelling acknowledge that learners' understanding of the correspondences between sounds and letters in the language influences their performance on literacy tasks, and must be accommodated within the models. There are two competing, though complementary, approaches to classifying these correspondences, based on large- or small-grain subunits. Generally, syllables or rimes define the large-grain approach, and phonemes the small grain (see Ziegler & Goswami, 2005).

Frith's (1985) three-phase theory of literacy acquisition suggests that each phase is associated with a particular acquisition strategy, which in turn is influenced by a particular grain size for the subunits: logographic, alphabetic, and orthographic. The large-grain first phase permits the instant recognition of familiar words; the second phase uses small-grain knowledge of correspondences between individual phonemes and graphemes, initially in spelling; and, the final phase involves operating on larger units than the alphabetic phase, forming something similar to a limited set syllabary. Phase 2 is the only one that requires a connection to phonological processes, and deficits at this point are seen as being very detrimental to the successful acquisition of decoding strategies. Reading failure is viewed as an artefact of alphabetic writing systems in this scheme, which can be demonstrated as specific impairments on non-reading tasks that involve phonological processing. It is also linked to the orthographic transparency of the language, with dyslexic and non-dyslexic German-speaking children making fewer errors on words and non-words than English children (Frith *et al.*, 1998; Landerl *et al.*, 1997), a key feature being the difference in vowel consistency for the two languages.

The four-phase model proposed by Ehri (1995) extended Frith's small-grain alphabetic phase into a partial and a mature phase. The partial phase is characterized by first and last letter cues for pronunciation, which is differentiated from the mature alphabetic phase that places greater emphasis on the formation of connections between phonemes and graphemes, and especially the influence of phoneme–grapheme (spelling) correspondences in reading development. The final phases for the two models are very similar, with recurring letter patterns becoming consolidated. This applies not only to the 70 letters or letter combinations that symbolize phonemes, but also to common large grain rime stems (Ehri, 1997), linking this phase to connectionist models of word reading (Ehri & Soffer, 1999).

Seymour (1999) criticized the Frith model because it does not show causal influences and dependencies, and modified his own dual foundation model (Seymour, 1997) to provide a four-phase model that better expresses the temporal dimension. A pre-literacy phase is introduced for epilinguistic language organization, with children showing sensitivity to rhymes, but unable to explain what a rhyme is. This is followed by the dual foundation phase, consisting of logographic and alphabetic aspects that develop according to teaching methods. Ehri's (1992) emphasis on small-grain GPCs is acknowledged by the suggestion that a high criterial level of letter-sound knowledge (80%) is necessary before these dual processes can advance. Compton (2000) has extended this idea, showing that although letter-sound knowledge is essential for early literacy growth in first-grade children (6- to 7-year-olds), knowledge of larger graphemes (consonant blends and digraphs, and vowel digraphs) is a more powerful predictor of non-word reading growth rate.

Seymour's (1999) later orthographic and morphographic phases move towards large-grain abstract representations of the spelling system, which are also acknowledged to parallel the phonological and semantic networks in connectionist models (Plaut, McClelland, Seidenberg, & Patterson, 1996).

The knowledge that phonemic segments of speech are not usually available to early readers has also led to the development of large grain theories for early reading acquisition, suggesting that initially phonological contributions may be at some higher level than the phoneme. Treiman (1987) suggested a progressive theory of phonological development from spoken words to syllables, to the segmentation of syllables into onsets and rimes, and finally onsets and rimes into phonemes. This was developed by Goswami (1993) who placed rime and analogy in the initial phase of reading English, associated with the analytic phonics approach to literacy acquisition. Goswami (1986) demonstrated that children in their second year of schooling can use analogies to facilitate the reading of unfamiliar words, and suggests that children should be trained in phonological skills (Goswami, 1999), focusing on rimes and onsets. This approach has also been linked to orthographic transparency when comparing German and English children (Wimmer & Goswami, 1994). English pupils were seen as using a phonological underpinning of direct access strategies that is predominantly at the onset-rime level. However, this was not a strategy found to be present in a later study in which English-speaking children did not show great proficiency in using onsets and rimes (Frith *et al.*, 1998).

Although the starting-point for these developmental theories varies, with the approaches differentially emphasizing the logographic, phonemic, and onset/rime aspects, they all indicate a progressive understanding of written language that ends with knowledge of the functioning of increasingly larger grain units, and there is increasing acceptance of this developmental trend, from serial to parallel processing, in children (Bijeljac-Babic, Millogo, Farioli, & Grainger, 2004).

While developmental models of reading are concerned with accounts of word acquisition rates and accuracy/error rates that may help formulate pedagogical approaches to literacy, computer models tend to be more concerned with accounting for adult processes, especially on lexical decision tasks and speeded word naming. Accuracy rates and word latencies have been shown to be substantial correlated ($r = -.81$) in a recent study of 40,481 words (New, Ferrand, Pallier, & Brysbaert, 2006), suggesting that accuracy/error rates in young children may translate into differential response times in adults. Frith *et al.* (1998) demonstrated lower error rates and faster response times for comparisons of German and English children across the age range 7–12 years, for words and non-words. Similar results were obtained for dyslexic children aged 11–12 years (Landerl *et al.*, 1997), although higher error scores on low frequency words for English subjects led to few latency measures. These findings were also replicated when English, French, and Italian young adult dyslexics were compared (Paulesu *et al.*, 2001), with the Italian group showing faster response times and significantly fewer errors for both words and non-words when compared with the two opaque language groups.

Although not concerned directly with developmental aspects, computer models were originally driven by simulations of acquisition processes using training sets of words. The parallel processing connectionist triangle (CT) model still derives its settings from the large grain 'properties of the training set, including the frequencies of words and the structural relationships among them (e.g. the extent to which they share patterns that are pronounced similarly or differently)' (Seidenberg & Plaut, 1998, p. 234). In this model, the ease with which a word is pronounced depends on its relative orthographic body transparency. Consistent words have body letter patterns that have the same phonological rime (e.g. *need,* since *-eed* → /iːd/ in all words containing the body). Inconsistent words, on the other hand, are those where the body is contained in words with different phonological rimes (e.g. *head*, since *-ead* → /iːd/ in some words and /ed/ in others). In this scheme, a simple, dichotomous classification may be applied to the transparency of individual words (consistent/inconsistent), but a more sophisticated categorization that refers to both regularity and consistency (see Jared, 2002) may also be appropriate, including a continuous representation of consistency that allows examination of transparency effects in a more detailed manner (Treiman, Mullenix, Bijeljac-Babic, & Richmond-Welty, 1995). Plaut (1999) dismisses the dichotomous distinction, claiming that 'language knowledge is inherently graded, and the language mechanism is a learning device that gradually picks up on statistical structure among written and spoken words and the contexts in which they occur' (Plaut, 1999, p. 544).

Serial processing models, such as the dual-route cascaded (DRC) model are driven by small grain GPC rules. Early versions of the DRC model claimed that the GPC rules, which translate letter strings into strings of phonemes, were not specified *a priori*, but were automatically learned, from exposure to the spellings and pronunciations of real words (Coltheart, Curtis, Atkins, & Haller, 1993). But for the DRC this has changed over time, and it now features hard-wired GPC rules no longer derived from a learning algorithm (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), and with this, there is an explicit admission that the DRC model does not offer any account of how reading is learned (Coltheart, 2006). Perry, Ziegler, and Zorzi (2007) see this absence of learning to be a major shortcoming of the DRC model because it cannot be used to simulate reading development or developmental reading disorders. Their connectionist dual process (CDP) model is claimed to be superior because its 'delta rule' learning has been widely applied to human learning, and as such has a psychological reality. It is also claimed to be superior to both the DRC and CT models because it is highly sensitive to the graded statistical consistency of spelling–sound relationships at multiple grain sizes (from letters to word bodies). In other words, the CDP model has a stronger developmental strand than alternative computer models, operating with both small- and large-grain subunits, and therefore can account for reading development in different languages when pre-trained on GPCs that reflect teaching strategies. Hutzler, Ziegler, Perry, Wimmer, and Zorzi (2004) demonstrated the need to link computer models to developmental data when investigating simulations of acquisition rates in German and English computer networks. Their initial network accuracy simulations showed the opposite pattern to children's cross-language learning rates, with small differences in the early phases and large differences later in the process. However, when the network was pre-trained on GPC rules taken from phonics programmes for young children in both languages, prior to training on words, the results showed a large initial advantage in the more transparent language, with a 25% decrease over time, capturing the essence of the children's cross-language learning rate effect.

A major challenge for computer models has been their inability to satisfactorily predict naming latencies on the individual items on which they are trained. Spieler and Balota (1997) suggested that the amount of variance predicted by successful computer models should be at least as strong as the strongest single factor correlating with human performance, and should be similar to that derived from factors typically shown to be involved in reading. They found that the simple predictors, log frequency, orthographic length, and neighbourhood density (Coltheart's *N*) collectively accounted for 21.7% of the item variance, whereas the CT model only accounted for 3.3%. Coltheart *et al.* (2001) found that the DRC model accounted for 3.49% and the CDP model for 7.73% of variance for the same database. This contrasts with the log frequency factor that alone accounted for 7.1% of the item variance. More recently, Perry *et al.* (2007) have demonstrated that a revised version of the CDP model, the CDP +, accounts for variance that is close to the three simple factors (17.3%).

Spieler and Balota (1997) were also concerned that the complexity of the associations between the continuous factors (e.g. frequency, regularity) made factorial experimental design both difficult and potentially misleading, and that models that capture the continuous nature of such factors are more likely to offer insights into the processes of skilled reading and, hence, their acquisition. They have recently developed this approach, using regression techniques, and for naming tasks with adults, four lexical factors (length, frequency, neighbourhood size, and consistency) were found to account for 15–17% of the variance, at the item level (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004).

For many years, it was assumed that orthography–phonology relationships operated on word recognition in a feedforward manner. However, Stone, Vanhoy, and Van Orden (1997) challenged this assumption, reporting that both feedback and feedforward consistency influenced printed word recognition (at least for English). Kessler, Treiman, and Mullenix (2007) reviewed the evidence for this claim and concluded that there was not convincing evidence for feedback consistency effects. They suggest that such results are surprising because 'inconsistency in the sound-to-letter direction, something that might logically make writing difficult, would seem to play no necessary role in reading, which involves mapping letters to sounds' (p. 159). However, Balota *et al.* (2004) view their feedback rime consistency effects as compatible with a highly interactive system in which both spelling and reading patterns contribute to the naming process as it unfolds across time.

The orthographic depth of a language affects the strategies used to teach foundation literacy skills. For shallow orthographies with consistent GPCs, there is often an emphasis on synthetic phonic methods (Seymour *et al.*, 2003). This approach (Johnston & Watson, 2004) has proved to be successful for English, although the mixed methods advocated by Duncan, Seymour, and Hill (1997) may be more appropriate for opaque languages ‘in which commonly occurring words contain letter structures which are inconsistent with the principles of simple grapheme–phoneme correspondence’ (Seymour *et al.*, 2003, p. 166). Goswami (2005), reviewing synthetic phonics, also points out that there are so many irregular words in English that it ‘may be difficult or impossible to access meaning by recoding letters into phonemes’ (p. 277). Instead, Ziegler and Goswami (2005) argue for the teaching of large-grain onset-rime units because they are more consistent for English. However, they acknowledge that this analytic phonics approach entails a greatly increased number of units being learned.

It appears from this that there is an emerging convergence of computer and developmental models that lend support to regression model outcomes identifying the major lexical factors involved in the acquisition and practice of literacy skills in opaque languages. The present study investigates the impact on children’s single word reading accuracy of those factors (length, frequency, transparency) that have been shown to influence adult reading performance.

Method

Participants

Data were collected for all 6-year-old pupils (Year 1, the second year of reading instruction) who in June/July 2003, 2004, 2005, and 2006 attended an urban Hull, UK, primary school. The academic performance of the children at the school is in line with national average levels. Over the 2003–06 period, 79.9% of the pupils attained level 4 or above in national tests in English, Maths, and Science, compared with the national average of 79.4%. The school uses a mixture of synthetic phonics and the National Literacy Strategy ‘Progression in phonics’ (Department of Education and Employment, 2000) approach to foundation literacy. A 16-week programme of synthetic phonics (Johnston & Watson, 1997) is introduced in the first term of schooling, and this is followed by the Oxford Reading Tree scheme, which is the adopted reading scheme, for use throughout the school. The same teacher taught the children for the duration of the research using the same approach for all groups.

The proportion of girls and boys, age, and standardized scores on the National Foundation for Educational Research (NFER) Group Reading Test II, Sentence Completion Form A [6–14] (NFER, 2000), a reading test commonly used in UK schools, are presented in Table 1. The test was administered to the pupils, as a group, by their class teacher in early June of each year, as part of the monitoring process for children at the school. The class teacher also scored the completed tests and determined the standardized scores. ANOVA results revealed no statistically significant differences between the 4 year groups for the standardized reading scores ($F(3, 101) = 0.36$, $MSE = 221.26$, $p = .78$).

Materials and procedure

Letter-sound knowledge for the lower-case 26 letters of the alphabet, presented in random order for each pupil, was tested on a Compaq Evo N1020v computer, with 15 in. display. Practice trials were not given. Each letter was presented, in the centre of

Table 1. Pupil assessment scores for years 2003–06

	2003		2004		2005		2006		2003–06	
Female	13		13		15		15		56	
Male	11		14		12		12		49	
N	24		27		27		27		105	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age	6.38	0.31	6.39	0.27	6.35	0.31	6.30	0.33	6.35	0.30
NFER	108.38	14.93	108.26	13.15	101.41	13.95	104.85	17.79	105.65	15.12
Letter sound	24.71	1.30	23.56	3.45	24.33	2.86	22.78	3.83	23.81	3.10
Advanced grapheme	7.33	3.23	6.48	4.14	5.70	4.15	8.26	6.90	6.93	4.87
Word reading	91.63	39.79	78.70	43.40	69.74	43.50	80.59	60.90	79.84	47.75

Note. NFER, NFER Group Reading Test II, Sentence Completion Form A [6–14], standardized score; letter sound, letter-sound knowledge (maximum 26); advanced grapheme, advanced grapheme knowledge (maximum 18); word reading, word reading accuracy [arcsine transformed] (maximum 150).

the screen, as a black letter on white background in the primary font, 72 point, that was designed to match the fonts used in the reading scheme. There was no time constraint in any of the tests, and new items were only presented after the pupil had made a response. If no response was forthcoming, the pupil was asked if they were ready for the next item, which was then presented. Pupils were asked to give the 'letter sounds', a term they were familiar with from their phonics programmes of work. The presentation of the sequence was advanced by the researcher clicking either the left or right mouse button. A left click identified the response as correct, a right click as incorrect, and each click recorded the accuracy of the pupil's response in the individual pupil file held on the computer. Only letter sounds were accepted as correct responses. If letter names were given, the pupil was asked to provide the sound. Feedback was not given to the pupils.

In addition to letter-sound knowledge for the 26 letters of the alphabet, advanced grapheme knowledge was tested for the following letter combinations: ai, ar, ch, ear, ee, er, ie, ng, oa, oi, oo, or, ou, ow, oy, sh, th, ue. These letter combinations were chosen because they represent the graphemes associated with phonemes in the phonics programme used in the school. The computer presentation procedure was the same as for the letter-sound test, with pupils being asked to give 'the sounds made by the letters'.

Following this, the reading accuracy for 150 high frequency words was tested, with all words presented in lower-case. The stimuli were the most frequent printed words in English according to the Lancaster–Oslo–Bergen Corpus (Hofland & Johansson, 1982) and were selected because their early acquisition should form the basis for effective later reading; they represent 50% of the total adult token count. The words were presented in random order for each pupil, in two blocks of 75 items, with a 30 min break between blocks. The first block of items was drawn at random from the full set of 150 words. The second block was randomly drawn from the remaining 75 words. In cases of doubtful pronunciation, pupils were asked to put the word into a sentence. The computer presentation procedure was the same as for the letter-sound test.

Measures employed in the study

The study aimed to examine the influence of a range of word characteristics on the reading accuracy of the children in the sample. The variables included were word frequency, word length, graphemic complexity, GP, PG, and SG probability, and orthographic neighbourhood density (*N*). Descriptions of these variables follow.

Word frequency

Balota, Pilotti, and Cortese (2001) warned that frequency counts may vary widely depending on the source materials. This variability may be as large as 10% in the variance accounted for by available word-frequency estimates (Balota *et al.*, 2004) in naming and lexical decision tasks. Spencer (2007), using spelling to dictation data from UK children aged 6–11, demonstrated that the predictive power of word frequency counts was influenced by factors such as the age range covered by the source materials and where they were sampled (e.g. USA vs. UK). For the present study, the frequency values from the children's printed word database (CPWD; Masterson, Stuart, Dixon, & Lovejoy, 2003) were the most appropriate for the children sampled, since the corpus involves a representative sample of UK books for children in the first 4 years of school. Values aggregated across the reading schemes were available, as were values for the Oxford Reading Tree scheme used in the school. The $\log_{10}$ transformed Year 1 frequency values from the CPWD were used in the analyses.

Graphemic and phonemic word length and phonetic difference

Spencer (2007) found that the difference between the number of graphemes and phonemes, the phonetic difference (PhD) that reflects the grapheme complexity of a word, was a powerful predictor of word-spelling difficulty. When orthographic length was analysed in terms of phoneme length and grapheme complexity, complexity was also the more powerful factor in predicting the reading difficulty of English words for young children (Spencer, 2001). However, it should be noted that because $\text{word letter length} = \text{phoneme length} + \text{complexity}$, only the latter two variables were entered into the analysis.

Transparency measures

An initial decision when considering how word metrics may be calculated concerns the corpus of words that will form the basis for the values. Hanna *et al.* (1966) used a dictionary made up of 17,310 word types, and this corpus formed the basis for the correspondences derived by Berndt *et al.* (1987). Spencer (2007) demonstrated that transparency measures from a smaller corpus of the 3,500 most frequent adult words were better predictors of pupil performance than those obtained from larger corpuses, confirming McGuinness's (1997) concerns about distortions within larger corpora.

The present study derived new metrics, at the small grain grapheme–phoneme level, from the most frequent 1,000 printed words in Masterson *et al.*'s (2003) Year 1 word frequency database, providing a close match between source and participants. When abbreviations and abbreviated forms (e.g. Mr, Mrs) were removed the number of words was reduced to 971.

For the grapheme–phoneme grain size, two parsing processes were applied to the 971 words to derive word metrics: the words were first divided into phonemes and then into the corresponding graphemes. The on-line version of the Oxford English Dictionary (2006) to original was used to provide phonemic definitions. The alignment of graphemes with phonemes has no standardized procedure, and the present study used a computer program that followed the principles established by Gontijo *et al.* (2003) for parsing word letters into graphemes. This process gave 44 phonemes and 121 graphemes, with 217 combinations of phonemes and graphemes (sonographs) for the 971 word corpus (Spencer, 2009).

This division of the 971 words into phonemes and graphemes allowed three basic metrics to be calculated:

(1) Sonograph frequency probability for each of the 217 sonographs (SG value, as derived by Spencer, 2009).
(2) Phoneme–grapheme correspondence probabilities of the 217 sonographs for each of the 44 phonemes (PG value, as derived by Hanna *et al.*, 1966).
(3) GPC probabilities of the 217 sonographs for each of the 121 graphemes (GP value, as derived by Berndt *et al.*, 1987; Gontijo *et al.*, 2003).

Each metric was calculated with (token) and without (type) weighting for individual word frequency. In addition to values derived from the 971 words in the children's printed word database, the data from Gontijo *et al.* (2003), Hanna *et al.* (1966), and Spencer (2007) were re-analysed to provide SG, PG, and GP metrics. Type values for the Gontijo *et al.* data could not be calculated because only token data were available. Token weighting was not calculated for the Hanna *et al.* data.

The metrics derived from the four sources were used to calculate transparency values for each of the 150 words in the experimental stimulus set used for reading aloud with the children. Two transparency values for each word were obtained from each of the three transparency metrics (see Table 2 for an example). The mean word transparency has been found to be a less powerful predictor of word reading and spelling than the least transparent phoneme (LTP) value, which is obtained by identifying the component with the smallest probability (Spencer, 2001, 2007).

Table 2. Example of mean and least transparent phoneme characteristics

P		G	PGP	GPP	SGP
s	–	s	.8028	.6098	.0453
e	–	ai	.0185	.1429	.0005
d	–	d	.9126	.9882	.0432
	MWT		.5779	.5803	.0297
	LTP		.0185	.1429	.0005

Note. P, phoneme; G, grapheme; PGP, phoneme–grapheme (spelling) probability; GPP, grapheme–phoneme (reading) probability; SGP, sonograph probability; MWT, mean word transparency; LTP, least transparent phoneme.

Orthographic neighbourhood size

Orthographic neighbourhood size was originally defined by Coltheart, Davelaar, Jonasson, and Besner (1977) as a measure of the degree to which a word is similar to other words, and represents the number of real words that differ by a single letter from the word. They found that for adults it had no effect on a lexical decision task for words. However, when the same words were used as stimuli for children, Laxon, Coltheart, and Keating (1988) found that items with many neighbours, which they termed 'friendly' items, proved significantly less difficult than those with few friends, in lexical decision and naming of words and non-words. This has been confirmed by Calhoon and Leslie (2002) who also found that words with large neighbourhoods are easier to read for beginning readers. Orthographic neighbourhood size for the 150 experimental words was obtained from the Masterson *et al.* (2003) database, and from the English Lexicon Project database (Balota *et al.*, 2002).

Results

Mean scores for the 105 Year 1 participants on letter-sound knowledge, advanced grapheme knowledge, and single word decoding of 150 common words are given in Table 1, according to year group. The results revealed no statistically significant differences between the groups for letter-sound knowledge ($F(3, 101) = 2.05$, $MSE = 9.31$, $p = .11$), advanced grapheme knowledge ($F(3, 101) = 1.39$, $MSE = 23.47$, $p = .25$), word reading accuracy ($F(3, 101) = 0.89$, $MSE = 2287.158$, $p = .45$), or standardized reading scores ($F(3, 101) = 0.36$, $MSE = 221.26$, $p = .78$). On this basis, the data from the 4 years were pooled for further analyses.

For the analyses, the proportion of children correctly reading each word was calculated. These item totals were submitted to arcsine transformation. The influence of the experimental variables on the transformed reading accuracy data were explored using a series of regression analyses. Prior to this, correlational analyses were used to explore relationships within the data.

Correlational analyses

Although Compton (2000) found that knowledge of letter sounds predicted reading growth rate in 6- to 7-year-old children, knowledge of complex graphemes was a more powerful predictor. In the present study, with children of the same age, knowledge of letter sounds is significantly correlated with both word reading accuracy ($r = .39$) for the 150 words in the study and the standardized reading scores ($r = .45$). However, as shown in Table 3, knowledge of complex graphemes has a stronger association with both ($r = .85$; $r = .79$), confirming Compton's results.

Table 3. Correlations between letter-sound knowledge, complex grapheme knowledge, word reading accuracy for 150 common words, and standardized reading scores for Year 1 pupils (pooled for 2003–06)

	1	2	3	4
1. Letter sound	–	.35**	.39**	.45**
2. Advanced grapheme		–	.85**	.79**
3. Word reading			–	.92**
4. NFER				–

Note. Letter sound, letter-sound knowledge (maximum 26); advanced grapheme, advanced grapheme knowledge (maximum 18); word reading, word reading accuracy [arcsine transformed] (maximum 150); NFER, NFER Group Reading Test II, Sentence Completion Form A [6–14] standardized score; $^{**}p < .01$.

The correlations between the transformed word accuracy scores and the word characteristics are shown in Table 4. The range of significant correlations between word characteristics and word reading accuracy is from $-.69$ to .98. Only five metrics do not

Table 4. Correlations between word reading scores for 150 common words for Year 1 pupils 2003–06, and word metrics (LTP values)

		Transparency																								
	Source	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1. Y1R		–	.48**	.46**	.49**	.59**	.57**	.48**	.45**	.45**	.51**	.58**	.51**	.44**	−.03	−.09	.16	.08	.02	.18*	.77**	−.69**	−.36**	−.59**	.51**	.47**
2. SG Ty	SC		–	.89**	.79**	.78**	.75**	.80**	.65**	.58**	.54**	.59**	.55**	.51**	.23**	.21*	.34**	.17	.15	.24**	.19	−.31**	.01	−.51**	.25**	.27**
3. SG Ty	SA			–	.91**	.75**	.76**	.81**	.53**	.52**	.51**	.51**	.49**	.45**	.23**	.24**	.33**	.18*	.15	.19	.18	−.33**	−.04	−.56**	.22**	.23**
4. SG Ty	H				–	.72**	.74**	.75**	.52**	.54**	.59**	.55**	.55**	.49**	.08	.09	.30**	.06	.02	.23	.23**	−.41**	−.07	−.65**	.31**	.28**
5. SG To	SC					–	.98**	.90**	.73**	.71**	.70**	.80**	.76**	.67**	.11	.06	.22**	.15	.10	.27**	.40**	−.43**	−.11	−.63**	.30**	.29**
6. SG To	SA						–	.92**	.69**	.69**	.67**	.75**	.74**	.66**	.09	.06	.19*	.12	.09	.24	.40**	−.42**	−.12	−.61**	.26**	.25**
7. SG To	G							–	.65**	.63**	.61**	.66**	.65**	.68**	.14	.10	.19*	.13	.11	.28**	.30**	−.35**	−.08	−.55**	.19*	.20*
8. PG Ty	SC								–	.94**	.84**	.85**	.84**	.81**	.20*	.10	.27**	.15	.07	.10	.19	−.34**	.01	−.49**	.31**	.38**
9. PG Ty	SA									–	.90**	.85**	.89**	.84**	.11	.02	.17*	.08	−.01	.07	.21*	−.36**	−.02	−.48**	.30**	.34**
10. PG Ty	H										–	.85**	.88**	.78**	−.05	−.11	.12	−.04	−.12	.03	.26**	−.44**	−.07	−.55**	.38**	.33**
11. PG To	SC											–	.94**	.80**	.06	−.05	.16*	.13	.01	.17	.39**	−.52**	−.17*	−.56**	.42**	.41**
12. PG To	SA												–	.83**	.02	−.06	.07	.07	.00	.11	.34**	−.46**	−.14	−.52**	.37**	.35**
13. PG To	G													–	.03	−.07	.09	.03	−.03	.21	.26**	−.42**	−.08	−.51**	.29**	.34**
14. GP Ty	SC														–	.91**	.56**	.87**	.83**	.15	−.05	.07	−.07	.23*	.12	.19
15. GP Ty	SA															–	.60**	.78**	.83**	.18	−.11	.12	−.02	.25*	.10	.16
16. GP Ty	H																–	.46**	.46**	.31**	.01	−.21*	−.14	−.16	.36**	.35**
17. GP To	SC																	–	.95**	.26**	.14	−.03	−.19*	.15	.19*	.24**
18. GP To	SA																		–	.31**	.07	.03	−.17	.24*	.16	.20*
19. GP To	G																			–	.12	−.27**	−.16	−.26**	.26**	.26**
20. Freq	M																				–	−.56**	−.44**	−.34**	.40**	.39**
21. WL																						–	.67**	.72**	−.68**	−.61**
22. PhL																							–	.00	−.50**	−.42**
23. PhD																								–	−.46**	−.45**
24. N	M																								–	.86**
25. N	ELP																									–

Note. Pearson correlations, except Spearman for PhD. Y1R, reading score (transformed); SG, sonograph probability; PG, phoneme–grapheme probability; GP, grapheme–phoneme probability; Freq, log word frequency; WL, word letter (orthographic) length; PhL, word phoneme (phonemic) length; PhD, phonetic difference; *N*, orthographic neighbourhood size; M, Ty, type; To, token; Masterson *et al.* (2003); SA, Spencer (2009), Adult's Corpus A 3K; SC, Spencer (2009), children's corpus C 1K; H, Hanna *et al.* (1966); G, Gontijo *et al.* (2003); ELP, English Lexicon Project, Balota *et al.* (2004); $^{**}p < .01$, $^{*}p < .05$.

significantly correlate with word reading, and these are the feedforward GP measures that intuitively would be expected to be associated with reading; the remaining GP measure has only a weak association. The sonograph and spelling measures all have high, significant correlations with the reading accuracy measure. The mean correlations between word reading and continuous word transparency metrics derived from the four corpora are: sonograph $r = .51$; PG (spelling) $r = .49$; GP (reading) $r = .05$. The correlations between reading accuracy and the remaining predictor variables are all significant, with frequency ($r = .77$), word length ($r = -.69$), and complexity ($r_s = -.59$) showing particularly strong associations.

The correlations between transparency metrics derived from the four corpora are generally high and significant. The mean correlations within type and token counts for the three metrics are: SG, $r = .82$; PG, $r = .86$; GP, $r = .56$. The correlations between metrics are more variable, the mean values are: SG/PG, $r = .61$; SG/GP, $r = .17$; PG/GP, $r = .06$.

The word transparency metrics that are weighted for word frequency values (token counts) tend to have larger associations with word reading than unweighted measures, but this will tend to reduce their power as predictor variables in the regression analyses because they are compromised by their significant associations with word frequency. Word frequency is also significantly correlated with word letter length ($r = -.56$), phoneme length ($r = -.44$), complexity ($r_s = -.34$), and the two values for Coltheart's N ($r = .40$; $.39$).

Regression analyses

The present study takes the approach recommended by Cohen, Cohen, West, and Aiken (2003) and Tabachnick and Fidell (2001) for hierarchical multiple regression analyses. The independent variables are entered sequentially to evaluate each variable in terms of what it adds to the prediction that was different from that afforded by variables entered earlier in the sequence. Variables are entered in varying permutations, which has the advantage of studying models that predict similar proportions of the total variance, but from different combinations of variables. This is a compromise between simultaneous entry and exploratory stepwise analyses. For each analysis, plots were obtained for residuals and indicated that the assumptions of linearity and homoscedasiticity were met, and that there were no outliers.

Given the restricted range of the complexity (PhD) values and small number of words within the PhD range 3–4, this variable was entered into the regression analysis as a series of dummy variables. This technique allows for the introduction of categorical variables as independent variables in multiple regression analysis, in this case adopted to prevent biasing from the few most complex words.[1]

---

[1] *In Table 4, non-parametric Spearman correlations are provided for the complexity (PhD) variable. For the dummy variables in the regression analysis, the reference group, the least complex group of words that had PhD values of 0, was chosen to reflect the recommendation of Cohen* et al. *(2003) that it 'should serve as a useful comparison . . . and should not have a very small sample size' (p. 303). The reference group for coding is assigned a value of 0 for every code variable, whereas the other groups are coded 0 or 1. For example, the coding scheme for the complexity variable allows for four categories: PhD = 0, 1, 2, 3, or 4. A word that has PhD = 0 would be coded across the four categories 0000, whereas a word that had a PhD = 4 would be coded 0001. It is through the use of dummy codes that categorical information can be rendered into quantitative form. Taken as a set, the number of categories − 1 (g-1) represents all the information contained in the nominal categories, and for this reason any category may act as the reference group. In the regression analysis g-1 categories are entered: PhD = 1, 2, 3, or 4.*

Spieler and Balota (1997) demonstrated that connectionist computer models capture at best slightly more variance than simple log frequency models, and substantially less than the combined predictive power of log frequency, neighbourhood density, and orthographic length. Balota *et al.* (2004) argued that the search for significant effects will not typically motivate researchers to report the unique amount of variance that a given factor accounts for in a design, but that it should be the driving force in the literature because the contribution that a variable makes, assessed by its standardized beta value and unique variance, indicates its power when comparing models. With this in mind a series of regression analyses was undertaken starting from a single variable model, with unique variances of each additional variable calculated.

Table 5 shows a hierarchical model for predicting children's word reading accuracy, with four steps. Steps 1–3 assess the variance attributable to the main variables identified in the literature, word frequency and word length. Log frequency has the highest association with reading accuracy (Table 4) and is entered in Step 1. This single variable model accounts for .60 of the variance in reading accuracy for the 6-year-old

Table 5. Summary of hierarchical regression analyses for variables predicting Year 1 word reading accuracy [arcsine transformation] ($N = 150$)

Source	Predictor variable	*B*	*SEB*	β	R^2	ΔR^2
Step 1						
M	Freq	0.34	0.02	0.77***	.600	
Step 2						
	PhL	−0.01	0.02	−0.03	.600	.001
Step 3						
	PhD1	−0.13	0.03	−0.26***		
	PhD2	−0.24	0.03	−0.35***	.725	.125***
	PhD3/4	−0.31	0.05	−0.28***		
Step 4a						
SC	SG	6.77	1.29	0.24	.770	.044***
Step 4b						
M	N	0.00	0.00	0.08	.729	.003
ELP	N	0.00	0.00	0.05	.727	.002
SC	SG	6.77	1.29	0.24***	.770	.044***
SA	SG	6.82	1.40	0.23***	.765	.039***
H	SG	5.53	1.30	0.22***	.756	.031***
SC	PG	0.18	0.04	0.19***	.754	.028***
SA	PG	0.16	0.04	0.19***	.752	.026***
H	PG	0.15	0.04	0.19***	.750	.025***
SC	GP	0.06	0.04	0.07	.730	.005
SA	GP	0.05	0.04	0.07	.730	.004
H	GP	0.08	0.04	0.09*	.733	.008*

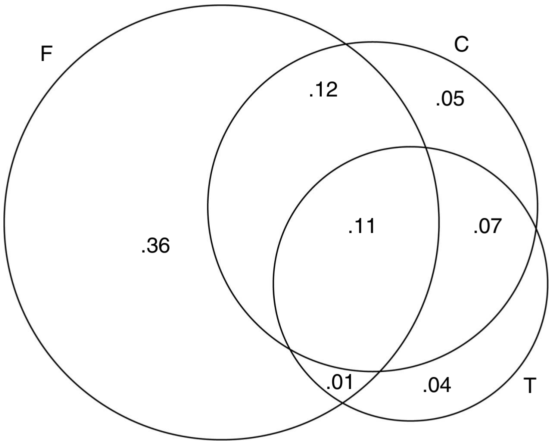
Note. SG, sonograph probability, LTP, type; PG, phoneme–grapheme probability, LTP, type; GP, grapheme–phoneme probability, LTP, type; Freq, log word frequency; WL, word letter length; PhL, word phoneme length; PhD, phonetic difference; *N*, orthographic neighbourhood size; *M*, Masterson *et al.* (2003); SA, Spencer (2009), Adult's Corpus A 3K; SC, Spencer (2009), children's corpus C 1K; H, Hanna *et al.* (1966); ELP, English Lexicon Project, Balota *et al.* (2004); *** $p < .001$, * $p < .05$.

children. Word letter length is entered in two separate steps by decomposing it into its two components, phonemic length and complexity. In Step 2, phoneme length is entered, but fails to reach significance, adding only .001 to the variance accounted for by the model. In Step 3, complexity is entered, as three dichotomous categorical variables, all of which are significant, adding .125 to the variance accounted for by the model. Overall this model, which is based on frequency and word length, accounts for .725 of the variance in reading accuracy.

Steps 4a and 4b assess the impact of transparency and orthographic neighbourhood size variables. At Step 4a all transparency and *N* metrics in Table 3 were entered into an exploratory stepwise regression, to determine the best statistical fit following Steps 1–3. The results show that the least transparent type value, derived from the UK children's database, has the greatest impact, adding .044 to the model total variance (.770). For this group of young children this demonstrates that sonograph measures are more powerful predictors of reading accuracy than spelling or reading correspondence values, or orthographic neighbourhood estimates. However, stepwise regression may select variables for inclusion in the model on the basis of only slight differences in their semi-partial correlations (Field, 2005), and will eliminate variables that could have very similar effects in the model. For example, although the transparency value derived from the children's database has been selected for inclusion in the model at Step 4a, it is reasonable to ask to what extent this model differs from one with the transparency values derived from UK adult sources, or US adult sources. Step 4b provides answers to these questions. Rather than applying the Step 4a stepwise procedure after Steps 1–3, each of the transparency and *N* variables is entered individually at Step 4b: Steps 1–3 are run, and at Step 4b a single variable is entered; this is repeated for all 20 transparency and *N* variables, with only a single variable being entered each time. The results of this procedure demonstrate that least transparent measures are more powerful predictors than mean transparency values, and that type values are more powerful than token metrics. In the interests of brevity, only the results for least transparent type variables are summarized in Table 5, Step 4b. The mean individual variances for the three transparency (type LTP values) metric groups and *N* are: SG $sr^2 = .038$; PG $sr^2 = .027$; GP $sr^2 = .006$; N $sr^2 = .003$. These results show that models that predict substantial proportions of reading accuracy variance in young children may be constructed from frequency and orthographic length variables, with significant additional transparency contributions from all sonograph probabilities and phoneme–grapheme (spelling) correspondence probabilities, but not from grapheme–phoneme (reading) probabilities, or neighbourhood metrics.

The hierarchical regression procedures adopted for this study, allowing variables to be entered in a systematic manner that is dictated by theoretical matters, offer an advantage over simple stepwise methods. The advantage over simultaneous entry methods is that they are not compromised by highly intercorrelated variables that result in unacceptable multicolinearity. However, a disadvantage is that because variables are entered in a particular sequence, the individual contributions of variables entered early in the procedure may be inflated. In order to look at the complexities of the interactions between variables a further series of analyses are necessary for the final model. In this analysis, all combinations of the variables are entered in a series regression analyses. For example, the final model in Step 4b, with transparency represented by the type value for the least transparent sonograph, based on the children's corpus, is essentially a three variable model (phonemic length does not make a substantial contribution to the model). This requires seven separate analyses to determine all individual variances and covariances that are shown in Figure 1. Here, it can be seen that the individual variances, with frequency partialled out, for complexity (.05) and transparency (.04) are similar, with a shared variance of .07.

The above results suggest that transparency measures should also be correlated with reading accuracy for word components, such as complex graphemes. The complex grapheme test was designed to measure reading accuracy for items that are taught in phonics programmes in UK schools, and that were used to train the CDP model (Hutzler *et al.*, 2004). Table 6 shows the correlations between accuracy scores for the 17 digraphs in the test and their transparency values for the pronunciations defined in the phonics scheme used in the school. Both sonograph (.65) and phoneme–grapheme (spelling) correspondence (.50) measures are significantly correlated with complex grapheme reading scores, but GPC probabilities are not. This reflects the findings for the word reading accuracy measures.



Source	Predictor variable		B	SE B	β	t	sr^2
M	Freq	F	0.28	0.02	0.64	***	.36
	PhD1		−0.09	0.03	−0.17	***	
	PhD2	C	−0.16	0.04	−0.24	***	.05
	PhD3/4		−0.21	0.05	−0.19	***	
SC	SG	T	6.58	1.31	0.23	***	.04

*** p < 0.001

Figure 1. Proportional Venn diagram of three variable model [$R^2 = .76$, $F(5, 144) = 91.05$, $MSE = 0.016$, $p < .001$]. $F =$ log frequency (Freq), based on Masterson *et al.* (2003) [M]; $C =$ complexity (PhD 1/2/3-4), $T =$ transparency (SG), based on Spencer (2009), children's corpus C 1K [SC].

Table 6. Correlations between complex grapheme reading accuracy [arcsine transformation] and transparency metrics ($N = 17$)

	1	2	3	4
1. CG	–	0.65**	0.50*	0.22
2. SG		–	0.66**	0.38
3. PG			–	0.27
4. GP				–

Note. CG, complex grapheme reading accuracy score [arcsine transformed]; SG, sonograph probability, LTP, type; PG, phoneme–grapheme probability, LTP, type; GP, grapheme–phoneme probability, LTP, type; $^{**}p < .01$, $^{*}p < .05$.

Discussion

Seidenberg, Petersen, MacDonald, and Plaut (1996) commented on the high intercorrelations between word characteristics that are undermining attempts to determine their unique effects in experimental work. Balota *et al.* (2004), when considering this problem, focused on the utility of regression studies in providing a complementary approach to standard factorial designs. The present study confirms these complex relationships, and uses regression analyses to investigate their combined and independent impact on single word decoding for the foundation literacy phase UK children in the study.

The results for these children confirm the impact of printed word frequency on visual word recognition, even across the restricted high frequency band studied, and indicate a substantial reduction in reading of one standard deviation for words two frequency quartiles apart. The three-variable model in Figure 1 predicts 77% of reading variance, and log frequency, based on the recent count of British English from Year 1 children's texts (Masterson *et al.*, 2003), uniquely accounts for 36%, with additional significant contributions from word complexity and transparency.

Implications for developmental and computer models of visual word recognition

Complexity, and SG and PG transparency have strong associations: words with greater complexity are less transparent. However, both have significant individual impact on word reading, whereas, for the children in this study, orthographic neighbourhood size measures do not.

Complexity

The results indicate that when orthographic length is decomposed into its two main constituents, complexity and phonemic length, the latter has no significant individual effect in the model, whereas complexity has a significant, powerful effect. This result supports the 'whammy effect' of complex graphemes, observed by Rastle and Coltheart (1998) with adult latencies. The children's accuracy data in this study suggest that the ontogenesis of this latency effect in adults can be observed in reading accuracy in the foundation phase of literacy.

Complexity effects have been simulated in the DRC-L model. For example, the second letter of the digraph < oo > is described as meeting with a hostile reception: 'an already-active phoneme in its set will exert inhibition upon it, thus slowing the rate at which its activation rises – perhaps even blocking its activation altogether.' (Rastle & Coltheart, 1998, p. 279). In a similar vein, Rey, Jacobs, Schmidt-Weigand, and Ziegler (1998) observed the intriguing facilitatory effect on word identification times for words having a greater number of phonemes. They found that holding letter-length constant and increasing phoneme-length effectively decreased complexity, which in turn led to shorter identification times. This suggests that grouping letters into graphemes requires additional processing time because, by holding length in letters constant, decreasing the number of phonemes amounts to increasing the average phoneme complexity.

Rastle and Coltheart (1998) interpret 'whammy' effects in terms of serial left-to-right parsing of print. This process slows down when complex graphemes are encountered by skilled readers, and in children it may severely disrupt the decoding process. The results of the present study confirm Laxon, Gallagher, and Masterson's (2002) observation that for young children's reading, although complexity is more pronounced in the case of vowels than consonants, 'complexity does matter, even when the complex consonants are very commonly used.' (p. 282). Compton (2000) also observed a complexity effect, with knowledge of complex graphemes predicting rate of growth in decoding skills, which was seen to support the developmental models of Ehri (1992) and Perfetti (1991). It also confirms Seymour's (1999) model of early reading in which children move from simple letter-sound correspondences to more complex connections that include multi-letter grapheme units. The present study takes this a stage further by considering the relative transparency of the complex graphemes presented in phonics schemes in the UK, and shows significant correlations between reading accuracy scores for complex graphemes and phoneme–grapheme (spelling) correspondences and sonograph probabilities. This illustrates the dilemma for the producers of such schemes because the convenient digraphs used to represent the vowels do not necessarily conform to the most frequent representations in the English orthography. For example, $<$ ie $>$ → /aI/ and $<$ ee $>$ → /iː/ are both used for teaching purposes (see Spencer, 2006), but the former is a very infrequent sonograph (SG = .0003) with only 14% of pupils providing the correct phoneme, whereas the latter is relatively common (SG = .0081) and is read by 64% of the pupils in this study.

Transparency

In all cases, mean transparency values for type counts showed smaller associations with word reading accuracy than the least transparent measures. Thus, 'tricky' sonographs can be seen to precipitate reading errors. The position of this sonograph in a word was entered into the regression analysis but did not make a significant individual contribution to the variance. The increased predictive power of least transparent measures, when compared with mean values, has previously been observed in both spelling and reading studies (Spencer, 1999, 2001, 2007)

All transparency metrics except, in general, GP correspondences have significant associations with word reading. The single GP value that has a statistical significant association with word reading is a token-based metric, and that association is not significant when frequency is partialled out. This presents problems for both dual route serial processing models and single route parallel models. The implicit assumption has always been that if reading is controlled by transparency measures, at any grain size, the direction of influence will be feedforward. The regression models presented in Table 4 may be interpreted as demonstrating that the learning process in children is rather different to that conceptualized by either of the main competing computer-based models. Spencer (2009) has demonstrated that GPC values are inherently unreliable metrics, showing that there is a substantial negative association between grapheme frequency and GPC metrics. In other words, there is a tendency for very infrequent graphemes to have high GPC values. For example, a grapheme that only appears once in a corpus of words from which metrics are derived will have a GPC value of 1.00, which is the same value as, or higher than, a very frequent and transparent grapheme. GPC metrics fail to discriminate between these poles, and actually suggest that a grapheme such as < eigh > → /eI/, which appears in only one word in the children's corpus (weight), should be read more easily than < t > → /t/. This anomaly was noted by Berndt *et al.* (1987) and also by Coltheart *et al.* (1993). The regression and correlation results suggest that for the children in this study, it is the frequency of sonographs in the corpus of words that they experience that influences the words they can read, rather than GP metrics. This offers a mechanism to explain the growth in knowledge about complex graphemes that are central to the developmental theories of Ehri (1997) and Seymour (1999), and the results of Compton's (2000) hierarchical linear models. Ehri and Soffer (1999) commented that figuring out the sonographs that make up the alphabetic system in English is not a straightforward process for children and that there have been few studies of the development of the awareness of these units. The present study suggests that for children who are taught within a 'phonics environment', as in the present study, the frequency of occurrence of the sonograph units in the children's reading materials has a substantial effect on the acquisition of such knowledge.

The results of the regression analyses may also point to a resolution of the feedback effects observed in adult latency studies. Feedback effects were not considered to be relevant in reading until the controversial study of Stone *et al.* (1997). Although studies have tended to focus on feedback rime consistency, Stone *et al.* indicate that this is only because it is a grain size that is very commonly used in studies, and feedback versus feedforward effects should not depend on choice of corresponding grain-size. Spencer (2001) demonstrated a small-grain feedback effect for Year 1 and 2 UK pupils, noting that the correspondence probabilities specifically derived for reading (Berndt *et al.*, 1987) were less effective predictors than the spelling values from which they were derived (Hanna *et al.*, 1966). Perry (2003) also found graded small-grain feedback effects for vowels. The results from the present study suggest that although the small-grain feedback effects appear to be counter-intuitive, they actually exert their influence through their strong association with sonograph frequency, which forms the basis for both feedforward and feedback measures.[2]

Orthographic neighbourhood size

Orthographic neighbourhood size, *N*, has received considerable attention in the literature, although its status is somewhat contradictory: it was originally anticipated that increased neighbourhood size would increase response latencies, according to contemporary models of lexical processing (Coltheart *et al.*, 1977), but instead

---

[2] *The strong association between the two measures reflects the constraint on the number of phonemes defining English. For phonemes, there is not the long tail of low frequencies that is obtained for graphemes. As the distribution of phonemes becomes more equal, so the association between PG and SG increases. Indeed, if all phonemes were present in equal numbers within the language, SG and PG values for words would be perfectly correlated. In this case, the feedback spelling effects would match the effects of SG frequency.*

a facilitatory effect was observed (Andrews, 1989, 1992; Laxon, Coltheart, & Keating, 1988). Reviewing the effects of neighbourhood size, Andrews (1997) has suggested that it simply acts as a surrogate for orthographic rime consistency, although Balota *et al.* (2004) have demonstrated an effect that is independent of feedforward and feedback rime consistency. Table 4 shows that although *N* has significant correlations with word difficulty, it is also associated with other word characteristics, especially word letter and phoneme length, frequency, and word transparency and complexity. Its contribution to the model in this study is not significant once word complexity and frequency have been partialled out of the regression, suggesting that in foundation literacy its effects may be because it acts as a surrogate for complexity.

Implications for cross-linguistic studies

There can be no doubt that learning to read in transparent orthographies is easier than in opaque languages such as English or French. However, defining the degree of transparency has not always been undertaken in a scientific manner. For example, Seymour *et al.* (2003) based their categorization on subjective judgments by the team members. Borgwaldt, Hellwig, and De Groot (2005) suggest that although there may be a general consensus on the approximate classification of the orthographic transparency of languages, there is relatively little quantitative cross-linguistic scientific research regarding it, with previous research dealing almost exclusively with monosyllables. They concluded that an analysis of this relatively small and biased subset might not always result in reliable assessments of orthographic transparency and proceeded to calculate entropy values for word-initial letter-to-phoneme correspondences on 10,000 to 120,000 word corpora for seven European languages. The assumption is clearly that large corpora are necessary to give an accurate description of the transparency of an orthography. However, the value of this approach is limited because, although the corpora are large, only the values for the initial letter-to-phoneme correspondences were calculated. Their entropy values do predict the rank order of children's accuracy scores for content words in Seymour *et al.*'s study for the five languages in common, and show significant correlations with adult naming latencies for three languages for which data was available (Italian, Dutch, and English). Although larger corpora may be viewed as offering greater accuracy when measuring orthographic transparency, McGuinness (1997) warned that they may distort values by including, for example, too many words borrowed from other languages, and recommended that corpora of 3–4,000 should adequately represent the metrics for the orthography of a language. A similar criticism can be made of measures using monosyllables extracted from extensive corpora containing unusual words of very low frequency. Spencer (2009) has demonstrated that sonograph probabilities from a small children's corpus are highly correlated with adult corpora of increasing size. As corpus size increases, additional sonographs only occur in the long tail of very low probabilities, the profile of the orthography remaining substantially the same. This has also been shown to be the case for the (reading) transparent Greek orthography (Spencer, Loizidou, & Masterson, in press) where a children's corpus of 4,162 words was represented by 100 sonographs, and the larger corpus of 217,664 (Protopapas & Vlahou, in press) was represented by an additional 18 low probability sonographs. In the present study, metrics based on the small children's corpus predicts a great proportion of the word accuracy variance than the larger adult corpora of Gontijo *et al.* (2003), Hanna *et al.* (1966), and Spencer (2007), suggesting that appropriate values for cross-linguistic comparisons of orthographic metrics may in future be computed from relatively small corpora.

The orthographic-depth hypothesis (Katz & Frost, 1992) suggests that opaque orthographies (i.e. deep in their terms) should promote access to stored lexical knowledge, and Lete, Pereman, and Fayol (2008) argue from this that the influence of word frequency on decoding accuracy as early as Grade 1 is a consequence of the high inconsistency of French orthography. This is also the case for both reading, in the present study, and spelling (Spencer, 2007) for Year 1 English children. On this basis, as transparency increases across languages, so the influence of word frequency may be expected to decline. This was observed by Frith *et al.* (1998) when comparing English and German, with only a small increase in reaction times from high- to low-frequency words for 8-year-old German-speaking children, and a significantly larger increase for the English children. A similar result was observed for accuracy, with little effect of frequency for German, but a strong effect for the English children.

The word length and complexity effects observed in this study suggest an initial disruption in young children's decoding in languages that appear to be transparent, but which have complexes of letter representing single phonemes. Reading and spelling are totally predictable for languages with single letter to phoneme orthographies and entropy values of zero (Finnish and Turkish). Oney and Goldman (1984) demonstrated that 7-year-old first-grade Turkish children's reading accuracy was substantially the same across words of 3 to 9 letters, unlike their American peers whose accuracy declined from 90 to 40% as letter length and complexity increased. They concluded that 'although irregular letter-sound correspondence may be restricted to only some patterns in a language, the effect is to increase the difficulty of acquiring decoding accuracy and speed for all patterns in a language.' (p. 564). Although Italian is seen as being a transparent language, it is not free from complexity in the form of two-letter graphemes, and is classed by Borgwaldt *et al.* (2005) as having an entropy value of 0.2. Cossu *et al.* (1995) investigated the effects of Italian orthographic complexity and found a significant word reading and spelling accuracy effect for first-grade children, which was exacerbated when the children were tested on long non-words, turning all complexities into 'an all most impossible task in both reading and spelling.' (p. 17). It is likely that other 'transparent' languages, such as Greek and Hungarian, will show complexity effects for reading and spelling in early school grades, and that for regular complexes this effect will fade for older children, as was observed by Oney and Goldman for their Grade 3 American children.

Conclusion

The results of this study offer a simple model of English literacy acquisition for young children, at the heart of which is frequency: the frequency with which a child encounters a word has long been known to affect the rate of acquisition of words; the frequency of individual word components, sonograph frequencies, are also shown to have a substantial independent effect in the present study. It is likely that these small grain effects will also be observed in languages that are also opaque for reading, such as French. For languages that are transparent for reading, but opaque for spelling, such as Greek, sonograph probabilities are also likely to have a greater effect on spelling performance than on reading.

The results support the view of Seymour, Bunce, and Evans (1992) who suggested that English literacy develops as a series of expansions from a core structure of simple vowels and consonants, to include common consonant complexes, and later, less common consonant graphemes, and complex vowels. They also complement Compton's (2000) view of the reciprocal nature of print knowledge and reading as a mechanism for this expansion: as children acquire more substantial knowledge of sonographs they can read more widely, and this brings them into contact with a wider array of increasingly more complex sonographs, the frequency of such encounters determining the extent to which they are internalized and made accessible for further successful decoding of whole words. This in turn demonstrates the extent to which the reading of English is dominated by complexity and frequency effects: token frequencies of individual words account for the largest proportion of word reading variance, with small-grain type frequency of individual sonographs adding significantly to this, and both interacting with grapheme complexity.

References

Andrews, S. (1989). Frequency and neighborhood size effects on lexical access: Activation or search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 802–814.

Andrews, S. (1992). Frequency and neighborhood size effects on lexical access: Lexical similarity or orthographic redundancy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 234–254.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin and Review*, *4*, 439–461.

Balota, D. A., Cortese, M. J., Hutchison, K. A., Neely, J. H., Nelson, D., Simpson, G. B., *et al.* (2002). *The English Lexicon Project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords*. St Louis, MO: Washington University, Retrieved from elexicon.wustl.edu

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*, 283–316.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory and Cognition*, *29*(4), 639–647.

Baron, J., & Strawson, C. (1976). Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 386–393.

Berndt, R. T., Reggia, J. A., & Mitchum, C. C. (1987). Empirically derived probabilities for grapheme-to-phoneme correspondences in English. *Behaviour Research Methods, Instruments, and Computers*, *19*, 1–9.

Bijeljac-Babic, R., Millogo, V., Farioli, F., & Grainger, J. (2004). A developmental investigation of word length effects in reading using a new on-line word identification paradigm. *Reading and Writing: An Interdisciplinary Journal*, *17*, 411–431.

Borgwaldt, S. R., Hellwig, F. M., & De Groot, A. M. (2005). Onset entropy matters – letter-to-phoneme mappings in seven languages. *Reading and Writing*, *18*, 211–229.

Calhoon, J. A., & Leslie, L. (2002). A longitudinal study of the effects of word frequency and rime-neighborhood size on beginning readers' rime reading accuracy in words and nonwords. *Journal of Literacy Research*, *34*(1), 39–58.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Hillsdale, NJ: Erlbaum.

Coltheart, M. (2006). Dual route and connectionist models of reading: An overview. *London Review of Education*, *4*, 5–17.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*(4), 589–608.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.

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.

Compton, D. L. (2000). Modeling the growth of decoding skills in first-grade children. *Scientific Studies of Reading*, *4*(3), 219–259.

Cossu, G., Gugliotta, M., & Marshall, J. (1995). Acquisition of reading and written spelling in a transparent orthography: Two non-parallel processes? *Reading and Writing: An Interdisciplinary Journal*, *7*, 9–22.

Cossu, G., Shankweiler, D., Liberman, I. Y., Katz, L., & Tola, G. (1988). Awareness of phonological segments and reading ability in Italian children. *Applied Psycholinguistics*, *9*, 1–16.

Department of Education and Employment (2000). *Progression in phonics*. London: DOEE.

Duncan, L. C., Seymour, P. H. K., & Hill, S. (1997). How important are rhyme and analogy in beginning reading? *Cognition*, *63*, 171–208.

Ehri, L. C. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In P. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 107–143). Hillsdale, NJ: Erlbaum.

Ehri, L. C. (1995). Phases of development in learning to read by sight. *Journal of Research in Reading*, *18*, 116–125.

Ehri, L. C. (1997). Learning to read and learning to spell are one and the same, almost. In C. A. Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to spell: Research, theory, and practice across languages* (pp. 237–269). Hillsdale, NJ: Erlbaum.

Ehri, L. C., & Soffer, A. G. (1999). Graphophonemic awareness: Development in elementary students. *Scientific Studies of Reading*, *3*, 1–30.

Ellis, N. C., Natsume, M., Stavropoulou, K., Hoxhallari, L., Van Daal, V. H. P., Polyzoe, N., *et al.* (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly*, *39*, 438–468.

Field, A. (2005). *Discovering statistics using SPSS*. London: Sage Publication.

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. E. Patterson, J. C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 301–330). London: Erlbaum.

Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German- and English-speaking children. *Scientific Studies of Reading*, *2*(1), 31–54.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 104–115.

Gontijo, P. F. D., Gontijo, I., & Shillcock, R. (2003). Grapheme–phoneme probabilities in British English. *Behavior Research Methods, Instruments, and Computers*, *35*(1), 136–157.

Goswami, U. (1986). Children's use of analogy in learning to read: A developmental study. *Journal of Experimental Child Psychology*, *42*, 73–83.

Goswami, U. (1993). Toward an interactive analogy model of reading development: Decoding vowel graphemes in beginning reading. *Journal of Experimental Child Psychology*, *56*, 443–475.

Goswami, U. (1999). Causal connections in beginning reading: The importance of rhyme. *Journal of Research in Reading*, *22*, 217–240.

Goswami, U. (2005). Synthetic phonics and learning to read: A cross-language perspective. *Educational Psychology in Practice*, *21*(4), 273–282.

Goswami, U., Porpodas, C., & Wheelwright, S. (1997). Children's orthographic representations in English and Greek. *European Journal of Psychology of Education*, 7(3), 273–290.

Hanna, P. R., Hanna, J. S., & Hodges, R. E. (1966). *Phoneme–grapheme correspondences as cues to spelling improvement*. Washington, DC: US Department of Health, Education and Welfare, [OE-32008].

Hino, Y., & Lupker, S. J. (2000). Effects of word frequency and spelling-to-sound regularity in naming with and without preceding lexical decision. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 166–183.

Hofland, K., & Johansson, S. (1982). *Word frequencies in British and American English*. Bergen: The Norwegian Computing Centre for the Humanities.

Hutzler, F., Ziegler, J. C., Perry, C., Wimmer, H., & Zorzi, M. (2004). Do current connectionist learning models account for reading development in different languages? *Cognition*, *91*, 273–296.

Jared, D. (2002). Spelling–sound consistency and regularity effects in word naming. *Journal of Memory and Language*, *46*, 723–750.

Johnston, R. S., & Watson, J. E. (1997). *Fast phonics first*. Spalefield: Spalefield Publishing.

Johnston, R. S., & Watson, J. E. (2004). Accelerating the development of reading, spelling and phoneme awareness skills in initial readers. *Reading and Writing: An Interdisciplinary Journal*, *17*, 327–357.

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 67–84). Amsterdam: Elsevier Science Publishers.

Kessler, B., Treiman, R., & Mullennix, J. (2007). Feedback consistency effects in single-word reading. In E. L. Grigorenko & A. J. Naples (Eds.), *Single-word reading: Behavioral and biological perspectives* (pp. 159–174). Mahwah, NJ: Erlbaum.

Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German–English comparison. *Cognition*, *63*, 315–334.

Laxon, V., Coltheart, V., & Keating, C. (1988). Children find friendly words friendly too: Words with many orthographic neighbours are easier to read and spell. *British Journal of Educational Psychology*, *58*, 103–119.

Laxon, V., Gallagher, A., & Masterson, J. (2002). The effects of familiarity, orthographic neighbourhood density, letter length and graphemic complexity on children's reading accuracy. *British Journal of Psychology*, *93*, 269–287.

Lete, B., Peereman, R., & Fayol, M. (2008). Consistency and word-frequency effects on spelling among first- to fifth-grade French children: A regression-based study. *Journal of Memory and Language*, *58*, 952–977.

Masterson, J., Stuart, M., Dixon, M., & Lovejoy, S. (2003). *Children's Printed Word Database. Economic and Social Research Council project (R00023406)*. Retrieved September 28, 2007, from http://www.essex.ac.uk/psychology/cpwd/

McGuinness, D. (1997). *Why children can't read and what we can do about it*. London: Penguin Books.

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin and Review*, *13*, 45–52.

NFER (2000). *Group reading test 6–12*. London: NFER-Nelson.

Oney, B., & Goldman, S. R. (1984). Decoding and comprehension skills in Turkish and English: Effects of regularity of grapheme–phoneme correspondences. *Journal of Educational Psychology*, *76*(4), 557–568.

Oxford University Press (2006). *The Oxford English dictionary on-line*. Oxford: Author, Retrieved June 1, 2006, from http://dictionary.oed.com/

Paulesu, E., Demonet, J.-F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., *et al.* (2001). Dyslexia: Cultural diversity and biological unity. *Science*, *291*, 2165–2167.

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S. F., *et al.* (2000). A cultural effect on brain function. *Nature Neuroscience*, *3*, 91–96.

Perfetti, C. A. (1991). Representations and awareness in the acquisition of reading competence. In L. Rieben & C. A. Perfetti (Eds.), *Learning to read: Basic research and its implications* (pp. 33–44). Hillsdale, NJ: Erlbaum.

Perry, C. (2003). A phoneme–grapheme feedback consistency effect. *Psychonomic Bulletin and Review*, *10*, 392–397.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP + model of reading aloud. *Psychological Review*, *114*, 273–315.

Plaut, D. C. (1999). A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. *Cognitive Science*, *23*, 543–568.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56–115.

Protopapas, A., & Vlahou, E. L. (in press). A comparative quantitative analysis of Greek orthographic transparency. *Behavior Research Methods*.

Rastle, K., & Coltheart, M. (1998). Whammies and double whammies: The effect of length on nonword reading. *Psychonomic Bulletin and Review*, *5*(2), 277–282.

Rey, A., Jacobs, A. M., Schmidt-Weigand, F., & Ziegler, J. C. (1998). A phoneme effect in visual word recognition. *Cognition*, *68*, B71–B80.

Seidenberg, M. S., Petersen, A. S., MacDonald, M. C., & Plaut, D. C. (1996). Pseudo-homophone effects and models of word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*(1), 48–62.

Seidenberg, M. S., & Plaut, D. C. (1998). Evaluating word-reading models at the item level: Matching the grain of theory and data. *Psychological Science*, *9*(3), 234–237.

Seymour, P. H. K. (1997). Foundations of orthographic development. In C. A. Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to spell: Research, theory and practice across languages* (pp. 319–337). Mahwah, NJ: Erlbaum.

Seymour, P. H. K. (1999). Cognitive architecture of early reading. In I. Lundberg, F. E. Tonnessen, & I. Austad (Eds.), *Dyslexia: Advances in theory and practice* (pp. 59–73). London: Kluwer.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*, 143–174.

Seymour, P. H. K., Bunce, F., & Evans, H. M (1992). A framework for orthographic remediation and assessment. In C. Sterling & C. Robson (Eds.), *Psychology, spelling and education* (pp. 224–247). Clevedon: Multilingual Matters.

Spencer, K. A. (1999). Predicting word-spelling difficulty in 7- to 11-year-olds. *Journal of Research in Reading*, *22*(3), 283–292.

Spencer, K. A. (2001). Differential effects of orthographic transparency on dyslexia: Word reading difficulty for common English words. *Dyslexia*, *7*, 217–228.

Spencer, K. A. (2006). Phonics self-teaching materials for foundation literacy. *Literacy*, *40*, 42–50.

Spencer, K. A. (2007). Predicting children's word-spelling difficulty for common English words from measures of orthographic transparency, phonemic and graphemic length and word frequency. *British Journal of Psychology*, *98*, 305–338.

Spencer, K. A. (2009). Feed-forward, -backward and neutral transparency measures for British English. *Behavior Research Methods*, *41*, 220–227.

Spencer, K. A., Loizidou, N., & Masterson, J. (in press). Feed-forward, -backward and neutral transparency measures for grade 1 and grade 2 Greek readers. *Behavior Research Methods*.

Spencer, L. H., & Hanley, J. R. (2003). Effects of orthographic transparency on reading and phoneme awareness in children learning to read in Wales. *British Journal of Psychology*, *94*, 1–28.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, *8*, 411–416.

Stanovich, K. E., & Bauer, D. W. (1978). Experiments on the spelling-to-sound regularity effect in word recognition. *Memory and Cognition*, *6*, 410–415.

Stone, G. O., Vanhoy, M., & Van Orden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language*, *36*, 337–359.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). New York: HarperCollins Publishers.

Treiman, R. (1987). On the relationship between phonological awareness and literacy. *Cahiers de Psychologie Cognitive*, *7*, 524–529.

Treiman, R., Mullenix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, *124*, 107–136.

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

Waters, G. S., & Seidenberg, M. S. (1985). Spelling–sound effects in reading: Time-course and decision criteria. *Memory and Cognition*, *13*, 557–572.

Wimmer, H., & Goswami, U. (1994). The influence of orthographic consistency on reading development: Word recognition in English and German children. *Cognition*, *51*, 91–103.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*, 3–29.

Ziegler, J. C., Jacobs, A. M., & Stone, G. O. (1996). Statistical analysis of the bidirectional inconsistency of spelling and sound in French. *Behavior Research Methods, Instruments, and Computers*, *28*, 504–515.

Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What is the pronunciation for -ough and the spelling for /u/? A database for computing feedforward and feedback consistency in English. *Behavior Research Methods, Instruments, and Computers*, *29*, 600–618.

Received 3 June 2008; revised version received 17 July 2009