



Word-effect variables in speeded-naming-task: Lexical and sub-lexical effects on reaction time and response duration in single-word-reading in an opaque orthography

Rosanna Emily Sparks

Word-effect variables in speeded-naming-task: Lexical and sub-lexical effects on reaction time and response duration in single-word-reading in an opaque orthography

### ABSTRACT

The study assessed the effects of word characteristics on monosyllabic word naming in English. 30 participants completed a speeded-word-naming task alongside two other tasks measuring word and non-word reading accuracy and fluency, and phonological awareness (TOWRE and Spoonerisms respectively). Responses were recorded (via the DMDX application) and reaction times (RT) and response durations (RD) were extracted by hand (using CheckVocal) and analysed. Multi-Level mixed-effects analyses indicated significant main effects on RT of word naming were: frequency and regularity. These findings were in line with previous research. The significant main effects on RD of word naming were: word frequency; item length; the orthographic similarity of words to other words; and bigram mean. Importantly, the study is the first of its kind to measure RD in word naming, in an opaque orthography. Interpretations of the results are discussed, including the possibility that the findings support a cascaded model of speech production, and suggest sub-lexical frequencies beyond the syllable. The double-deficit hypothesis of dyslexia is challenged, with RD findings suggesting that the factors influencing phonological awareness also influence rapid automatized naming (RAN). Limitations of the study and suggestions for future research are discussed.

KEY WORDS:	READING	OPAQUE ORTHOGRAPHY	WORD FREQUENCY	RESPONSE DURATION	CASCADED MODEL
------------	---------	-----------------------	-------------------	----------------------	-------------------

## Introduction

The ability to read is one of the most significant achievements of human civilisation and cognition, with word recognition being a focus of research since Cattell's (1886) innovative work. In fluent conversation two to three words per second are generated (Levelt, Roelofs & Meyer, 1999). Although one would assume word knowledge to be relatively simple due to an adult reader's vast knowledge of the skill, the processes involved are exceedingly complex. This complexity is highlighted by well over 100 years of experimental research devoted to understanding the processes involved in word recognition.

Balota and Yap (2006) discuss an often unstated assumption underlying models of visual word recognition. The assumption is that there is a 'magic moment' in word processing where the reader recognises the word, but as yet does not know its meaning. This seems reasonable as it would be difficult to interpret something until you have first recognised what that something is. The 'magic moment' is when lexical identification takes place. Once identification has occurred, a lexical representation is activated for a response to be executed, which in turn unlocks access to meaning. The two most common measures of the magic moment are lexical decision and speeded naming tasks. Both of these tasks can measure the reaction time (RT), that is, the time taken from stimulus onset to response onset.

The representational stages involved in lexical retrieval for the production of simple utterances have generated a great deal of interest in recent years (see Balota & Yap, 2006). Researchers are still debating the exact characteristics of these representations. However, it is generally accepted that lexical access in speaking can be subdivided into two stages, one that is concerned with the retrieval of semantic characteristics and one that involves access to the phonological properties of the intended words (see Levelt et al., 1999, for an overview). As the findings of the present study relate to current models of reading, these will briefly be discussed first, before returning to models of speech production.

### Computational models of reading

Although there are numerous computational models of reading, the two most influential approaches will be discussed. The Dual Route Cascaded model (DRC, Coltheart, Rastle, Perry, Langdon & Ziegler, 2001) assumes that there are two routes available to readers: a lexical route for reading words in which the lexical orthography is mapped to the lexical phonology; and a non-lexical route for reading non-words, which uses letter level spelling-sound correspondences, such as grapheme-phoneme correspondences or mapping rules, to read regular and non-words. The DRC model suggests that skilled readers utilise both the lexical and non-lexical routes. The DRC was built through effects of word characteristics (e.g. frequency) on the reaction times (RTs) of participants in tasks. However, the model does not account for any behaviour after the initial onset of an articulation, such as the effects of word characteristics on response duration (RD).

In contrast to the DRC, Triangle, or Parallel Distributed Processing (PDP, Seidenberg & McClelland, 1989; Plaut, McClelland, Seidenberg & Patterson, 1996; Harm & Seidenberg, 1999, 2004) models differ in several crucial ways. The PDP consists of two interacting subsystems: a phonological pathway which maps orthography to phonology and a semantic pathway linking phonological, semantic and orthographic units of representation. Unlike the DRC which assumes two separate routes, a lexical and a non-lexical pathway, for reading words and non-words, the PDP accomplishes both in a single system. This connectionist model of reading assumes that distributed representations and weighted connections between units are used when learning to read, rather than symbolic rules for mapping letters and sounds. Similarly to the DRC, the PDP does not account for any behaviour after the onset of an articulation (e.g., RD).

PDP models are intended to account for those factors that affect the computation of orthography to phonology. The models do not attempt to account for factors that affect the articulatory-motor component of naming performance (Spieler & Balota, 1997). In response to this, Seidenberg and Plaut (1998) state that human performance is affected by several factors beyond the scope of their model, including processes involved in recognising letters and producing articulatory output. Are these accounts correct to assume that RD is outside their remit? If speech production is syllabary - where syllables are taken to be the basic units of articulatory programming (Levelt et al., 1999) - then perhaps, yes. However, if speech production is shown to be cascaded, reading models must account for processes producing articulatory output, thus RD should be an important focus in models of reading as well as speech production.

A main question about lexical retrieval in speech production is how the two processes, the retrieval of semantic characteristics and access to the phonological properties of the intended word, relate to each other. Discrete two-step models (Levelt et al., 1999) assume that speaking proceeds in a serial manner from semantic to phonological retrieval, and that the two stages are independent from each other. With the semantic stage being complete before the phonological stage begins, there are for example, no semantic effects on phonological retrieval. Conversely, cascaded speech production models (Abrams & Balota, 1991; Kello, Plaut, & MacWhinney, 2000) dispute the modularity assumption and propose that processing proceeds gradually from one level to the other, and that semantic retrieval does not have to be fully concluded before phonological access begins.

Kello et al. (2000) reported an effect of Stroop interference on RDs, however, these only arose when there was increased pressure for speeded responding. This suggests that speech production could be both staged and cascaded, dependent on external factors. However, a cascaded model under certain settings can be seen to exhibit staged characteristics (see Damian, 2003). Hence, at the structural level, Kello et al.'s findings favour the cascaded view.

Damian (2003) found that RDs of target words were never systematically affected by semantic and phonological experimental manipulations, indicating that the two stages are independent from one another; providing support for the 'discrete' model of speech production. However, Meyer and Damien (2007) later reported that when a distracter picture was related to the target picture, response latencies were shorter than when the pictures were unrelated, concluding that their results were best explained within a cascaded model of lexical access.

The vast majority of research to date on the processes of reading and speech production, including the DRC and PDP models, and Levelt et al.'s (1999) influential work, have focused on processes before and up to the onset of an utterance (RT). There is very little research on the factors that influence RD in developmental reading. Some previous research findings offer ideas as to what can be expected, however. If target words are preceded by semantically related primes, it appears that the RD of the spoken responses may be shortened (see Kawamoto, Goeltz, Agbayani, & Groel, 1998). However, Kawamoto et al. recorded the onset durations (the initial part of the spoken response) and the rime durations (the last part of the spoken response) separately and therefore full RDs were not analysed. Additionally, Balota & Abrams (1995) reported that high frequency words, compared to low frequency words, influenced processes involved in word identification (RT), as the staged model assumes, but crucially that high frequency words also affected the execution of a response. However, the critical words were not spoken, and therefore, the measurement of RD was not that of the spoken utterance. For example, in three of their experiments the participants were asked to move a handle to either the left or right, depending on whether the word (as opposed to a pseudo-word) was on the left or right. The speed and force at which they moved the handle was the measurement. Balota and Abrams reported that the time taken to reach the high frequency words was shorter and executed with more force than that for low frequency words; suggesting that response preparation and response execution are not separate processes, but in fact do interact with each other.

As an articulatory response unfolds over time a potential measure is its duration, the time interval between onset and offset of a spoken utterance. Davies, Barbón & Cuetos (2013) looked at the performance of dyslexic and typically developing (TD) Spanish children in an oral reading task. They found that accuracy, RT and RD were affected by word frequency and length for both sets of participants. However, there are some drawbacks to the study. Most of the effects reported were significantly greater for the dyslexic group. It is also possible that the results were confounded as word characteristics, such as age of acquisition (AoA), were not controlled for whilst looking at word frequency. As these controls are crucial to the current study, the word-effect variables relevant to this study will be discussed next in detail.

### **Word-effect variables**

Frequency refers to the amount of times a word appears in print. It is a very common characteristic that is known to affect word processing times (Balota, Cortes, Sergent-

Marshall, Spieler, & Yap, 2004). In the DRC model lexical representations for high frequency words are activated more quickly (Coltheart et al., 2001), while in the PDP model high-frequency words have more heavily weighted connections through practice (Balota et al., 2006) and are thus activated quicker than low frequency words. However, neither the DRC or the PDP consider possible word effects on RD. Davies et al., (2013) found that more frequently encountered words were identified (RT) and spoken (RD) faster by Spanish children, than less frequently encountered words, although the RD affects are not explained in detail. This quicker identification may be linked to practice, because the more times an individual experiences a word the quicker their recognition of the stimuli will be. Previous findings (Balota & Abrams, 1995; Davies et al., 2013) suggest that the present study will find word frequency to have an impact on RT and RD of words, with more frequently encountered words having quicker RT and RD than less frequently encountered words.

It has been observed that the length of a word affects how quickly it is processed and identified, with increased word length requiring greater time for identification (e.g., New, Ferrand, Pallier, & Brysbaert 2006). However, it has been noted that more advanced readers use a more direct process of lexical access (Acha & Perea, 2008), and therefore adult readers generally only show length effects when reading non-words (Weekes, 1997). As the participants in the present study have all studied to degree level, and are therefore relatively advanced readers, it is not expected to find length effects on RTs of words. However, as longer words take longer to say, it is expected to find increased RDs for words with more letters than for those with fewer letters.

Orthographic neighbourhood ( $N$ ) refers to how similar in spelling words are to each other. This can be measured using Orthographic Levenshtein Distances (OLD), the minimum number of insertions, deletions or substitutions required to turn one word into another (Yarkoni, Balota & Yap, 2008). Andrews (1997) described how there are nearly always facilitatory effects of  $N$  on naming tasks. One might expect there to be a relationship with the length of a word and its neighbourhood size, as longer words are likely to have fewer neighbours. However, as the longest letter strings used in this study consisted of six letters, this relationship was not expected to be found. As current models of speech production stop at the onset of a word, it is uncertain how OLD will affect the RD of spoken words. If however, a neighbourhood size effect on RD is found, this would further suggest that the stages prior to articulation had not been concluded, thus potentially inconsistent with the modularity theory of speech production.

Age of acquisition (AoA) refers to when in an individual's life they learnt a word. Several studies have shown that words learned early in life are processed quicker than those learned later in life (e.g., Morrison & Ellis, 1995). Although new words are learned throughout life, those acquired earlier have a lasting advantage than more newly acquired words. Two possible reasons for this are: firstly, the concepts with

which earlier acquired words are learned are more deeply embedded in the networks of semantic knowledge (Steyvers & Tenenbaum, 2005) and therefore have stronger connections in tasks of verbal processing (Belke, Brysbaert, Meyer, & Ghyselinck, 2005). Secondly, it is believed that children have greater plasticity when learning words than do adults (Ellis & Lambon Ralph, 2000), and that this plasticity decreases with age resulting in later acquired words being at a disadvantage to those acquired earlier. It is therefore expected that AoA will affect the RT of words, but also if it is assumed that the stages are cascaded, it is possible that AoA effects will have an impact on the RD of words.

The imageability of a word refers to how easily a mental image of the word can be made. Strain, Petterson, and Seidenburg (1995) observed that words with high imageability (e.g., carrot) had faster RTs compared to those with low imageability ratings (e.g., scarce). Both the DRC and PDP models simulate facilitatory effects of imageability on RTs and therefore these effects are expected to be observed in the current study. Again, if imageability effects are seen on RDs of words, this may be suggestive of a cascaded approach in speech production.

### **The present study**

If the duration of an utterance remains fixed, and only a length effect is observed, in response to the manipulation of stimulus characteristics, then it can be concluded that these central processing stages were concluded before the response was initiated, and that articulation is a separate stage, supporting the idea of 'discrete' model of speech production. However, if articulatory responses are significantly lengthened or shortened due to the characteristics of the word (see word-effect variables above), this would imply that some extent of processing cascades to articulation, or is ongoing at the central level after response initiation has taken place; hence, the processing stages were not concluded before the response was initiated, providing support for a 'cascaded' model of speech production.

Drawing on previous research, it is suggested that the current study should find word characteristics effects in word naming in RDs as well as in RTs. This study will be the first to analyse both reaction time and response duration measures of individual's reading in an opaque orthography such as English. Analyses were designed to estimate the effect of item attributes on RT and RD, including the effects of lexical frequency, length, AoA, OLD and imageability.

### **Method**

#### **Participants**

30 participants (7 males and 23 females) ranging in age from 18 to 51 years (mean = 25.53, SD = 8.16) were recruited through opportunity sampling. The sample size was reflective of the constraints imposed by experimental time, as well as the lengthy preparatory process required in order to extract relevant data for analysis (approximately 90 minutes per participant). Participants had on average 16.73 years

of education ( $SD = 1.57$ ). All participants were native English speakers with normal or corrected to normal vision and reported no history of neurological illness.

### **Task descriptions**

The participants were tested individually in a quiet room, in one 35 minute session, in which they completed three tasks: a test of phonological awareness (Spoonerisms, Phonological Assessment Battery, Fredrickson, Frith, & Reason 1997), a standardised reading test (TOWRE, Torgessen, Wagner, Rashotte, Rose, Lindamood, Conway, et al., 1999) and a speeded-word-naming task. The tasks were presented in randomised order so as to negate order-effects. For the computerised test (speeded word naming), stimuli were presented and responses recorded on a Dell laptop computer (1024 x 600 (60 Hz) 32 bit RGB screen with a 16.70ms refresh rate), using the DMDX application (Forster & Forster, 2003). Participants were placed approximately 50cm from the screen.

### **Spoonerisms task**

The Spoonerisms task is part of the Phonological Assessment Battery (PhAB) (Frederickson et al. 1997) and assesses an individual's phonological awareness by measuring their ability to manipulate sounds. The spoonerisms test consists of two sections. Both sections contain 10 items. Both sections were preceded by an oral explanation and a practice session comprising of three examples. Each part had a time limit of 3 minutes. In the first section participants had to produce ten semi-spoonerisms, which involved replacing the first sound of a word with a new sound, (e.g., 'cot' with a /g/ gives 'got').

Participants were awarded 1 point for each correct response, 10 points in total were available for section one. In the second section subjects were asked to exchange initial sounds in two words, (e.g., 'daisy log' gives 'lazy dog') and thus producing two words per item, for ten pairs of words. For each correct word participants were awarded a point, therefore, 20 points were available for section two. Participant scores were marked out of 30 overall and all participants completed the same test. Participants were not shown the printed stimuli.

### **TOWRE**

The Test of Word reading Efficiency (TOWRE) was developed by Torgessen et al. (1999) and is a measure of word reading, accuracy and fluency.

The TOWRE contains two subtests: a test of Sight Word Efficiency (SWE) requiring word reading (104 items) and a test of Phonemic Decoding Efficiency (PDE) requiring non-word reading (63 items). Each subtest in the TOWRE contains practice items. Items on both subtests were presented in lists in columns on A4 sheets.

Participants were instructed to read the words aloud as quickly and accurately as possible, starting from the top left column and finishing at the bottom right. If a word could not be read, they were instructed to skip it and move on to the next. Subtest I, Sight Word Efficiency, was presented first to the participant and timing started at the



onset of the first word. For each subtest participants had 45 seconds, with their final score being the total number of words correctly read within this time. Subtest II, Phonemic Decoding Efficiency, was administered and scored in the same way.

### Word naming task

This test was used to analyse the reading performance of each participant in detail. 160 critical words and 10 practice words of high and low frequency were presented for reading aloud (see Appendix A).

170 items were selected varying on the following word characteristics: i.) word frequency: 170 items were taken from the LgSUBTLCD database of Brysbaert & New (2009). Adelman, Brown & Quesada (2006) observed that a measure of frequency should take into account contextual diversity, the number of contexts in which a word is seen, as opposed to the amount of times a word appears in print. Therefore, the measure of frequency used in this study is that put forward by Adelman et al. ii.) Word length, the amount of letters in a word. In the present study all words were monosyllabic and consisted of 3 to 6 letters. iii.) OLD, Yarkoni et al., (2008), describe OLD as the minimum number of insertions, deletions or substitutions required to turn one word into another. iv.) AoA: AoA ratings for all words are based upon the Kuperman norms (Kuperman, Stadthagen-Gonzalez & Brysbaer, 2013), which were derived from a study of 30,121 English nouns, verbs, and adjectives. v.) Imageability ratings were obtained from the Oxford Brookes University norms database (provided by supervisor) in which items were rated on a seven point scale from low to high imageability. vi.) BG\_mean (sub-lexical unit frequency) refers to average bigram frequency, which is the summed bigram frequency divided by the number of successive bigrams (Balota et al., 2007). A summary of the item characteristics are presented in Table 1.

**Table 1**  
**Descriptive statistics for item norm characteristics**

	<i>Mean</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
<b>LgSUBTLCD</b>	3.14	0.69	1.45	3.92
<b>Length</b>	4.30	0.73	3.00	6.00
<b>OLD</b>	1.51	0.28	1.00	2.05
<b>Age of Acquisition</b>	5.84	2.08	2.78	12.01
<b>Imageability</b>	4.36	1.28	1.66	6.76
<b>BG_Mean</b>	1675.86	813.85	168.33	4149.67

Participants were asked to read the letter-strings aloud as quickly and accurately as possible. Critical items were randomly assigned to five blocks of thirty-two words,

sixteen high frequency and sixteen low frequency words per block. Words were presented in random order per block, and blocks were presented in random order per session. The sequence of events in a trial was as follows: i.) A blank screen of 500ms ii.) An asterisk fixation mark in the centre of the screen for 500ms iii.) Presentation of the stimuli for 2000ms, the response interval. Words were presented on screen in 32-point Arial black lowercase letters at the centre of a grey field. Participants were asked to sit at a distance of approximately 50cm from the screen during testing.

Stimuli were presented and responses were recorded using the DMDX application. Participants wore a headphone set with a microphone attached, which was linked to the test computer to allow for DMDX to record the sound to the test computer hard drive directly, as .wav files.

### **Centering variables**

Multicollinearity is associated with both regression and word-effect analyses. It is a statistical phenomenon in which two or more predictor variables are highly correlated resulting in difficulty distinguishing the effects of the individual variables. To check whether the predictor variables of the current study were too correlated, correlation tables displaying pair-wise correlations for item and subject predictor variables were generated. The condition numbers for these subsets of predictors were then calculated based on methodology proposed by Baayen, Davidson and Bates (2008). Both the condition number for word norms (86.52) and the condition number for subject attributes (45.74) would be characterised as 'dangerous' by Baayen et al. (2008). To resolve this issue, predictor variables were centered on their means (Cohen, Cohen, West & Aiken 2003) by subtracting the mean values for each variable from the individual values of each variable, bringing variable values into a distribution around zero. Condition numbers were then recalculated (words norms=3.19; subject variables=2.39) with new figures reflecting safe levels of collinearity (Baayen et al., 2008).

## **Results**

### **Data extraction**

A total of 5100 responses were recorded. Of those, 300 were practice items and were not included in the analysis process. Sound spectrograms of the recorded responses were analysed by hand to extract RT and RD, using the CheckVocal application (Protopapas, 2007). Only correct responses were analysed (4727 responses), excluding all erroneous responses (73 responses, 1.54% - a very low error rate).

### **Subject attributes**

Statistics pertaining to subject attributes: age, years of education, TOWRE word and non-word scores and times, and spoonerisms scores are reported in Table 2. The

sample were well educated and performed near, but not at, the maximum level on all tests.

**Table 2**  
**Summary of demographic characteristics and standardised test scores**

	<i>Mean</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
<b>age</b>	25.53	8.16	18	51
<b>years of education</b>	16.73	1.57	14	22
<b>TOWRE words accuracy</b>	94.97	9.35	76	104
<b>TOWRE words time</b>	43.43	2.42	39	45
<b>TOWRE non-words accuracy</b>	58.17	5.43	45	63
<b>TOWRE non-words time</b>	42.13	4.32	30	45
<b>spoonerisms</b>	26.73	2.15	22	30

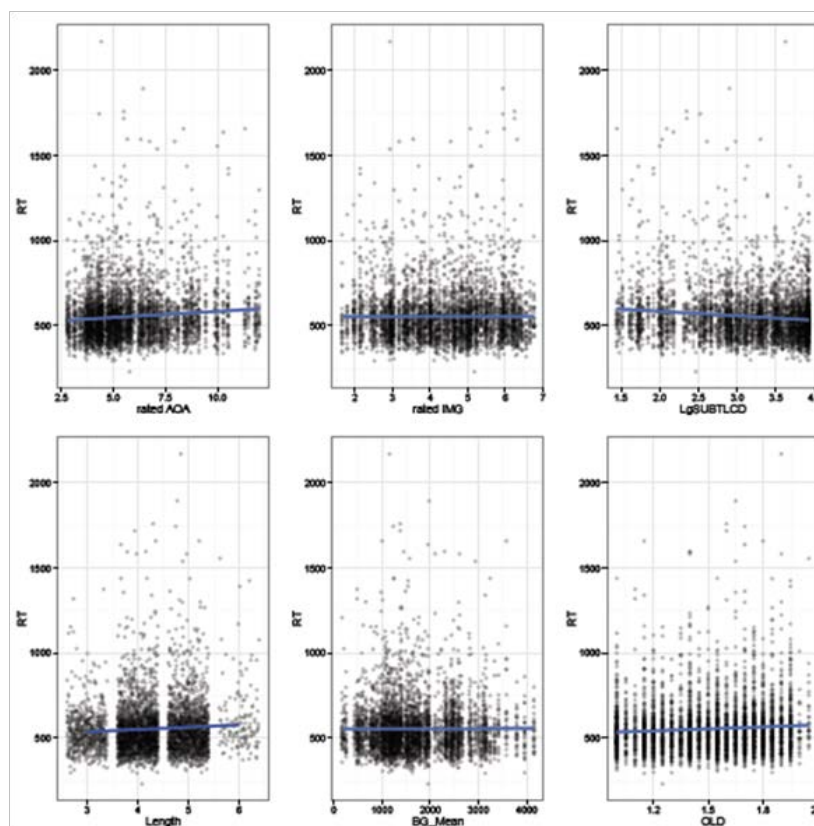
### **Mixed-effects modelling**

Traditional approaches to random effects modelling, such as ordinary least squares regression or univariate ANOVA, suffer multiple drawbacks. A main one being limited statistical power owing to problems caused by repeated observations (Baayen, Davidson & Bates, 2008); a drawback inherent to traditional word-effect analyses. The current study had a repeated measures design, with the same items presented to different participants, requiring the use of mixed-effects modelling. Mixed-effects models estimate both ‘fixed’ effects, for example, word frequency, and ‘random’ effects, for example, unexplained variation between subjects or items, while properly accounting for error variance (Baayen, 2008; Baayen, Davidson, & Bates, 2008) and thus eliminating the limited statistical power observed in the traditional approaches.

### **Mixed-effects modelling results for RT**

Following Davies, Barbon and Cuetos (2013), a series of models of increasing complexity were stepped through. Model comparisons were then conducted, using the Likelihood Ratio Test (LRT)  $-2\text{LogLikelihood}$  (2LL) (Pinheiro & Bates, 2000), to evaluate whether the added complexity of each successive model was justified by an improvement of fit. Model 0 was an empty model including just random effects of subjects and items but no fixed effects. The control variable, trial order, was added to make model 1, although the addition provided a non-significant LRT statistic ( $X^2 = 0$ , 1 df,  $p = 0.9963$ ). With the second model (model 2) phonetic coding was included and provided a significant LRT statistic ( $X^2 = 88.978$ , 10 df,  $p = 8.54e-15$ ), indicating that model 2 was justified. With model 3, main effects of subject attributes (centered variables of: participant’s age, TOWRE word score, TOWRE non-word score and Spoonerisms score) were added and again a significant LRT statistic ( $X^2 = 21.14$ , 4 df,  $p = 0.0003$ ) was observed, indicating that model 3 was a better fit to the data. Model 4 included the addition of main effects of item attributes (centered variables of: frequency, length, OLD, AoA, imageability and bigram mean) and again another

significant LRT statistic was yielded ( $X^2=79.251$ , 7 df,  $p= 1.958e-14$ ). Figure 1 shows the main effects of the raw data of item characteristics.



**Figure 1: Main effects of item characteristics (RTs in milliseconds)**

With the fifth model the subject effects were allowed to take on nonlinear rather than monotonic slopes. Baayen et al., (2006) document the advantages of regression with restricted cubic splines (RCS) to model nonlinearities. They highlight that serious consideration of non-linearities is an absolute prerequisite for accurate prediction. Therefore, models 5 and 6 use RCS. The comparison between model 4 and 5 yielded a non-significant result ( $X^2=1.0312$ , 4 df,  $p=0.905$ ) indicating that the addition was not justified by improved model fit to the data. In the sixth model item and subject effects took on nonlinear rather than monotonic slopes, again yielding a non-significant result ( $X^2=7.8485$ , 6 df,  $p=0.2494$ ). As models 5 and 6 were not justified, nonlinear item or subject effects were not included in subsequent analyses.

Next, model 4 was taken forward to check that the model effects were not influenced by the impact of non-significant effects. With model 7 all item and subject effects were included (all of which were linear because models 5 and 6 did not justify the inclusion of nonlinear item or subject effects). Model 7 was rerun but without its non-significant effects, to produce model 8. Model 8 illustrated that for all but the BG\_mean effect, effects that were significant if other factors were included were also significant if just the significant factors were included. Model 9 checked that the model outlier observations did not influence the pattern of results for model 8. Data

was subset to just observations not associated with large standardised model residuals. The results from models 8 and 9 suggest that all, but the BG\_mean effect, are effects that are not significant because they depend on the presence of outlier observations or non-significant other effects.

In the next phase of analysis, model 7 was taken forwards, holding the fixed effects constant between models but varying the composition of the set of random effects. This can be seen in model 10 which used the full dataset, rather than the outliers-removed dataset, and is reported in Appendix B.

The next set of models checked to see if random effects of subjects and items on intercepts and random effects of subjects on slopes of the significant item attribute effects - frequency, BG\_mean and regularity - were justified by improved fit to the data. The fixed effects were kept the same whilst varying the content in the random effects component of the model.

Model 10 was examined with both subject and item random effects on intercepts added to the model. To see if both random effects of subjects and items were justified, two models were run, each including just one of the random effects. Model 10S included just random effect of subjects and model 10I just random effect of items. The models were compared against model 10 with the -2LL test. Both were shown to be significant, model 10S,  $\chi^2 = 53.468$ , 1 df,  $p=2.629e-13$  and model 10I,  $\chi^2 = 2543.7$ , 1 df,  $p<2.2e-16$ . Therefore, random effects of both subjects and items were justified by significant improved model fit.

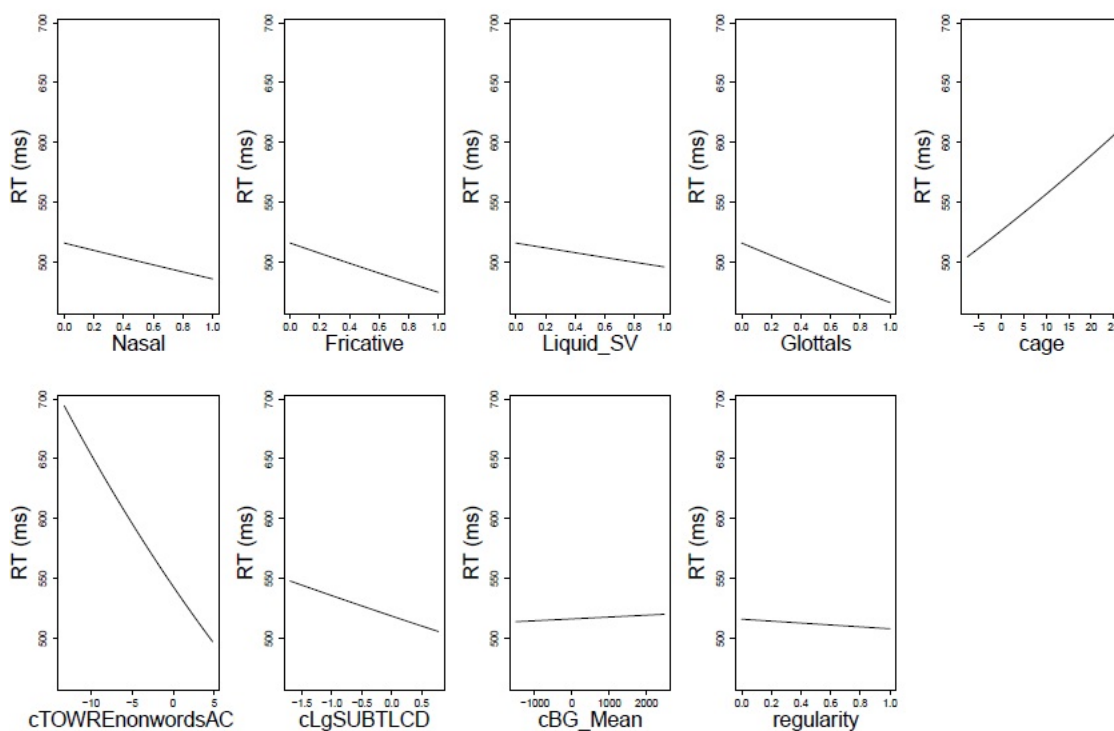
A series of models were then run examining if random effects of subjects on the slopes of the main fixed effects were justified by an improved fit to observed data. The random effect of subjects on slope of the frequency effect was added to the baseline model (model 10) producing model 10.a (significant,  $\chi^2 = 48.568$ , 1 df,  $p = 3.19e-12$ ). Then random effect of subjects on slope of the BG\_mean effect was added to make model 10.b (not significant.  $\chi^2 = 0.5055$ , 1 df,  $p=0.4771$ ). Lastly, random effect of subjects on slope of the regularity effect was added to give model 10.c (significant,  $\chi^2 = 11.236$ , 1 df,  $p=0.0008$ ). The frequency and regularity random slopes appeared justified, so model 10 was rerun including these additional terms to report if significant effects remained significant once random slopes were included (model 11). Model 11 suggested that the frequency and regularity effects did vary between participants but were reliable as overall effects.

Therefore, the final model for RT (model 12) was a model including only the significant effects, on the no-outliers database. Model 12 can be seen in Table 3, and Figure 2 shows the partial effects (model prediction of effects, taking into account other effects) plot for the final RT model.

**Table 3**  
**Summary of final model (12)**

Fixed effects	<i>Estimate</i>	<i>SE</i>	<i>t</i>	<i>pMCMC</i>
<b>(Intercept)</b>	2.7460	0.0106	258.84	0.0001
<b>Nasal</b>	-0.0259	0.0054	-4.81	0.0001
<b>Fricative</b>	-0.0360	0.0033	-10.97	0.0001
<b>Liquid_SV</b>	-0.0171	0.0045	-3.82	0.0002
<b>Glottals</b>	-0.0439	0.0055	-7.93	0.0001
<b>cake</b>	0.0024	0.0013	1.84	0.0084
<b>cTOWREnonwordsACC</b>	-0.0080	0.0020	-4.02	0.0001
<b>cLgSUBTLCD</b>	-0.0140	0.0020	-6.9	0.0001
<b>cBG_Mean</b>	0.0000	0.0000	0.75	0.4314
<b>regularity</b>	-0.0068	0.0027	-2.47	0.0098

**Note:** number of observations : 4727; 160 words; 30 participants

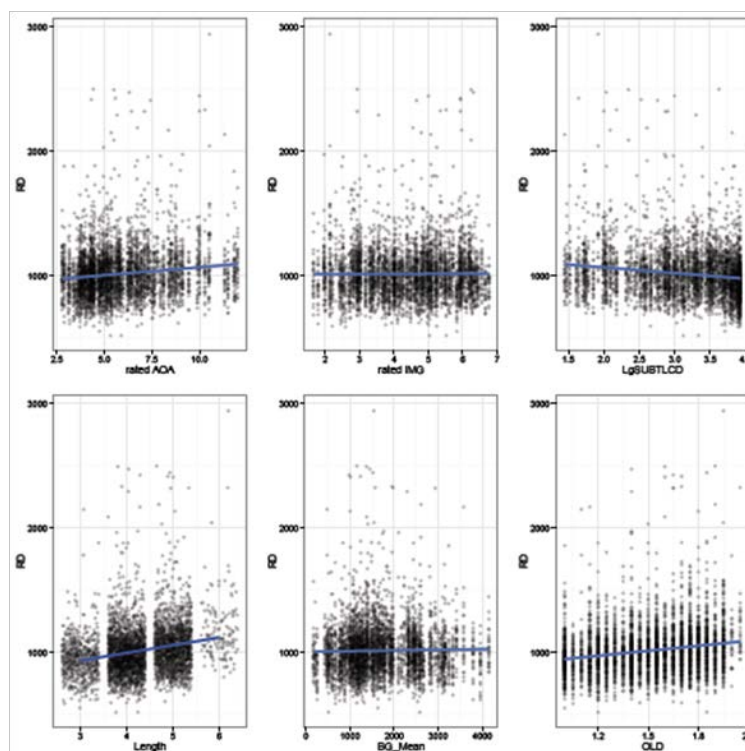


**Figure 2: RT partial effects plot for the final model**

**Note:** (c = centred, TOWREnonwordsACC = nonword reading accuracy, LgSUBTLCD = frequency)

### Mixed-effects modelling results for RD

The analysis process for RD mimicked that for RT. The first six models, each one increasing in complexity were the same as those for RT (see mixed-effects modelling results for RT). Using the Likelihood Ratio Test (LRT)  $-2\text{LogLikelihood}$  (2LL) (Pinheiro & Bates, 2000), models were compared to evaluate whether the added complexity of each successive model was justified by an improvement of the fit.



**Figure 3: Main effects of item characteristics (RDs in milliseconds)**

**Note:** Figure 3 illustrates the raw data, bivariate relations of the item characteristics: frequency, length, AoA, OLD, imageability and bigram mean.

Model 1 produced a significant LRT statistic ( $X^2=15.352$ , 1 df,  $p=8.923e-05$ ) when compared to model 0, therefore justifying the inclusion of the control variable, trial order. The addition of phonetic coding to model 2 provided a significant LRT statistic ( $X^2= 18.686$ , 10 df,  $p=0.0444$ ), indicating that model 2 was justified. Model 3, which included main effects of subject attributes produced a significant LRT statistic ( $X^2=11.844$ , 4 df,  $p=0.0186$ ), indicating that model 3 was a better fit to the data. Model 4 included the addition of main effects of item attributes and yielded another significant LRT statistic ( $X^2=88.357$ , 7 df,  $p= 2.69e-16$ ).

With the fifth model the subject effects took on nonlinear rather than monotonic slopes. The comparison between model 4 and 5 yielded a non-significant result ( $X^2=2.2538$ , 4 df,  $p=0.6892$ ) indicating that the addition was not justified by improved model fit to the data. However, in the sixth model where both item and subject effects took on nonlinear rather than monotonic slopes, a significant result was

yielded ( $\chi^2=18.754$ , 6 df,  $p=0.0046$ ). As model 6 was justified, nonlinear item effects were included in subsequent analyses.

Checks were then carried out to ensure that the model effects were not influenced by the impact of non-significant effects. With model 7 all item (nonlinear) and subject (linear) effects were included. Model 7 was rerun but without its specific non-significant effects, to produce model 8. In model 8 the AoA effect became non-significant while the frequency effect got stronger. It was also observed that the length effect was significant in the nonlinear component only. Model 9 checked that the model outlier observations did not influence the pattern of results for model 8. Data was subset to just observations not associated with large standardised model residuals. The results from models 8 and 9 suggest that all but AoA and non-linear length are effects that are not significant because they depend on the presence of outlier observations or non-significant other effects.

As with RT analysis, model 7 was taken forwards, holding the fixed effects constant between models but varying the composition of the set of random effects. This can be seen in model 10 which used the full dataset rather than the outliers-removed dataset and is reported in Appendix C.

Model 10 was examined with both subject and item random effects on intercepts added to the model individually. Model 10S including just random effect of subjects proved significant ( $\chi^2=470.26$ , 1 df,  $p<2.2e-16$ ) and model 10I just random effect of items also yielded a significant result ( $\chi^2=3076.9$ , 1 df,  $p<2.2e-16$ ). As seen with RT, random effects of both subjects and items were justified for RD by significant improved model fit.

A series of models were then run examining if random effects of subjects on the slopes of the main fixed effects were justified by an improved fit to observed data. Random effect of subjects on slope of the frequency effect were added to the baseline model (Model 10) producing model 10.a (significant,  $\chi^2=45.525$ , 1 df,  $p=1.507e-11$ ). Then random effect of subjects on slope of the OLD effect was added to make model 10.b (significant,  $\chi^2=7.1301$ , 1 df,  $p=0.0076$ ). Lastly, the random effect of subjects on slope of the BG\_mean effect was added to give model 10.c (not significant,  $\chi^2=0$ , 0 df,  $p=1$ ). The frequency and OLD random slopes appeared justified, so model 10 was rerun including these additional terms to report if significant effects remained significant once random slopes were included (model 11). Model 11 suggested that the frequency effect remained marginal whilst the OLD effect remained a reliable overall effect.

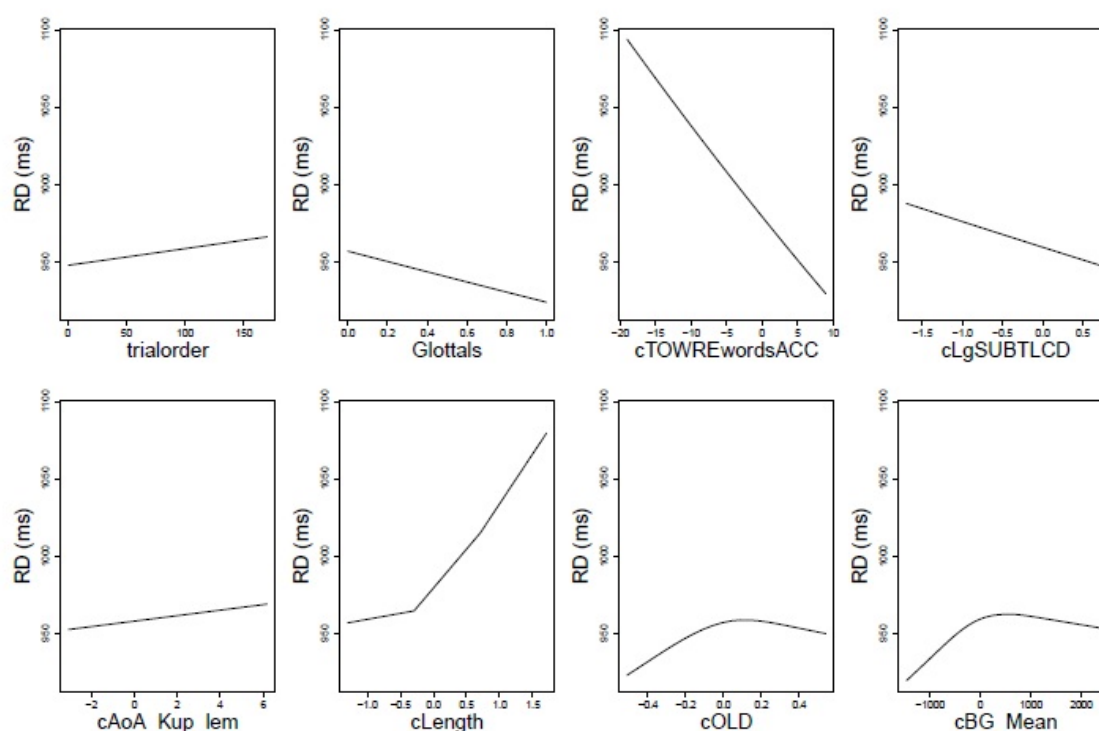
Therefore, the final model for RD (model 12) was a model including only the significant effects, on the no-outliers database. Model 12 can be seen in Table 4 and Figure 4 shows the partial effects plot for the final RD model.



**Table 4**  
**Summary of final RD model (12)**

Fixed effects	Estimate	SE	t	pMCMC
(Intercept)	2.9940	0.0114	262.36	0.0001
trialorder	4.9420E-05	1.1370E-05	4.35	0.0002
Glottals	-0.0153	0.0068	-2.25	0.0076
cTOWREwordsACC	-0.0025	0.0010	-2.58	0.0001
cLgSUBTLCD	-0.0074	0.0035	-2.14	0.0164
cAoA_Kup_lem	0.0008	0.0011	0.73	0.3874
r <sub>cs</sub> (cLength,3)cLength	-0.0012	0.0072	-0.16	0.8618
r <sub>cs</sub> (cLength,3)cLength'	0.0186	0.0071	2.63	0.0028
r <sub>cs</sub> (cOLD,3)cOLD	0.0375	0.0163	2.30	0.0094
r <sub>cs</sub> (cOLD,3)cOLD'	-0.0272	0.0172	-1.58	0.0728
r <sub>cs</sub> (cBG_Mean,3)cBG_Mean	1.4400E-05	6.2620E-06	2.30	0.0090
r <sub>cs</sub> (cBG_Mean,3)cBG_Mean'	-1.5580E-05	8.8700E-06	-1.76	0.0480

**Note:** Number of observations: 4727; 160 words; 30 participants  
 non linear=''  
 r<sub>cs</sub>= restricted cubic splines



**Figure 4: RD partial effects plot for the final model**

**Note:** (c = centred, TOWREwordsAC = word reading accuracy, LgSUBTLCD = frequency)

## Discussion

The present study investigated the effect of word and subject attributes on reading performance, as measured by the speeded-word-naming task, in an opaque orthography, English. The hypotheses were derived from current models of reading and speech, and from previous word-effect research. The significant main effects observed relating to faster RTs in the word-naming task were: words with +nasal, +fricative, and +liquid phonetic features in the initial (the beginning sound of a word); the older participants with better nonword (phonological coding) skills (as measured by the TOWRE nonwords test); and words that were higher frequency, and more regular. These findings were not surprising, providing similar results to previous research (e.g. Balota et al., 2004). As the results were consistent with the present understanding of RT in reading, they will not be discussed further.

The main significant effects observed relating to shorter RDs in the word-naming task were: words appearing earlier in the experiment (trial order effect); words starting with glottal; participants with better word reading skills (as measured by the TOWRE words test); words that were higher in frequency; shorter letter-strings; words with tighter orthographic neighbourhoods (words that require fewer insertions, deletions or substitutions to turn them into another word, see Yarkoni et al. 2008); and words with less frequent bigrams (BG). As the result of a more justified fit in the final model of analysis, the significant item effects in RD (frequency, length and OLD) were nonlinear.

The observation of these effects in RDs as well as RTs indicates that that lexical influence extends beyond the initiation of responses, consistent with the view that activation cascades through the phonological process in reading (Balota & Yap, 2006).

### Item attributes in RD

The RD effects of primary interest will be discussed in detail, with provided possibilities for what the findings may imply.

The length and BG mean effects reported were generally expected. Although reading models do not go beyond specification of phonemes, speech models do and from these, it can be inferred that length and BG mean effects reflect the articulatory programming demands for words; longer words take longer to say. A possible explanation as to why words with less frequent BG have shorter RDs is discussed below. However, a puzzling observation was that words with more common bits - bigram frequency reflects how common pairs of letters in words are - were not easier to say. The reason for this is unknown and may be an area of interest for future research.

The OLD effect is not accounted for by speech models and the observed effect was more surprising. Andrews (1997) suggests that neighbourhood/orthographic similarity effects reflect the impact of similarity on reading performance because of

the relation between  $N$  and the frequency of the parts of words; with words that are orthographically similar to many other words being recognised faster than more distinctive words. Of course it must be noted that all previous reading research has been concerned with RT or accuracy, but from the present findings it would not be unreasonable to postulate that OLD may reflect the frequency of parts of words and that words with more frequent parts are easier to say (RD) rather than prepare (RT).

It is possible that the OLD and BG mean effects, in this account, reflect differing levels of sub-lexical frequency. Levelt's speech model (Levelt et al., 1999; Levelt 2001) assumes the notion of a syllabary. The model suggests that a speaker has access to a repository of frequently used syllables of the language. According to the theory, the articulation of the word can be initiated as soon as all of its syllabic scores have been retrieved. It is important to note here that all words used in the speeded-naming-task were monosyllabic, so the OLD and BG mean effects can not be confused for syllable effects. The OLD effect illustrates that words with parts that look more like others, and have tighter orthographic neighbourhoods are easier to say (shorter RD) than more distinctive words with bigger orthographic neighbourhoods. The opposite effect, however, is observed for BG mean, with rare or more distinctive bigrams being easier to articulate, resulting in shorter response latencies. Why OLD is not competitive, but BG mean is cannot be answered with the present dataset. To clarify these findings it would be suggested to use priming in future research to confirm the effects observed. If the priming has an effect it can be inferred that OLD and BG mean do affect the pre-articulatory stage. The current findings, that the word-effect variables OLD and BG mean have a significant effect on RD, therefore suggest that perhaps sub-lexical frequency goes beyond Levelt's syllabary.

However, it is also possible that OLD actually reflects the impact of knowing a word as a whole object and how much that word is like other words, with those that are more like others being easier to say. This would connect with the observed effect of word frequency on RD.

The effect of frequency on RD might be interpreted as the result of a confound between the frequency of words and the frequency of the parts of words. This account would come from the theory in which RT represents a magic moment whereby the stimulus is perceived, semantic and phonological coding are retrieved, followed by phonetic coding and the onset of the response. Thereafter, everything else is articulation and motor programming. However, as the current study found length, OLD and BG mean effects, it would imply that a fair attempt to control for sub-lexical frequency effects has been made, and thus the effects are already taken into account.

The effect of frequency on RD might also be interpreted as owing to the way in which word sounds are stored. The Phonological Completeness Hypothesis of Brown and Watson (1987) proposes that the AoA effect emerges as a consequence of the

quality of an individual's phonological representations in the phonological output lexicon. It is proposed that those words learned earlier in life are stored as whole-word representations in the phonological lexicon. As a child's vocabulary increases however, the phonological store begins to represent words acquired later in a more segmented form. Therefore, it might be that more frequent/earlier acquired words are stored as whole sounds. Although this hypothesis was discredited by Monaghan and Ellis (2002), and although the theory is about RTs, the current data may provide findings in support of it. It is a possibility that the hypothesis could say something about how words are actually said given the significant effect of word frequency, and although not significant, words acquired earlier in life were spoken faster than those learned later.

However, the more interesting possibility is that the frequency effect on RD really is a word frequency effect. At minimum, this suggests that lexical processing extends beyond the initiation of responses, i.e., that reading processes are cascaded through the magic moment, consistent with the view of the Balota group (2004, 2006) but not explained by reading models at present.

It should also be mentioned that there is the possibility that the findings in the present study may be confounded with an unmeasured variable. There is no measure of syllable frequency in the model, which might propose a problem given the Levelt (2001) assignment of articulatory processing to a syllable-level set of mechanisms for motor programming. Additionally, although initial sounds of words were coded and controlled for, word endings were not. Future studies could code for word endings as well as initials. It is not anticipated that individual letters will affect speech production, however, to fully control for the structure of the word, the impact of individual letters could be tested by adding letter frequencies and comparing the response latencies.

### **Implications for future research**

If it is assumed that lexical reading processes are cascaded, as seen in the RD effects, the implications of these findings can significantly advance the literature on reading models, for example, the literature on rapid automatized naming (RAN, see Denckla & Rudel, 1976; Kirby, Georgiou & Martinussen, 2010; Manis, Doi & Bhadha, 2000). RAN refers to how quickly an individual can pronounce the names of a set of familiar stimuli. The four types of stimuli that have been used most often are colours, objects, numbers, and letters. There is a growing consensus that phonemic awareness, letter-sound knowledge and RAN are strong predictors of reading development in English (Caravolas, Lervåg, Mousikou, Efrim, Litavský, Onochie-Quintanilla et al., 2012). A slow naming speed is a characteristic of poor readers or those with dyslexia (Kirby et al., 2010). Wolf and Bowers (1999; see also Wolf, Bowers & Biddle, 2000) proposed the "double-deficit" hypothesis of dyslexia, in which reading deficits are more severe in individuals with weaknesses in both phonological awareness and RAN than in individuals with deficits in only one of these cognitive processing skills. Hulme and Snowling (2013) support the double-

deficit hypothesis by stating that RAN is statistically independent of letter knowledge and phoneme awareness as a predictor of reading development.

However, taking the current findings into account, RD effects may help to explain why RAN relates to reading development. It can be speculated that RAN is related to reading development because lexical reading processes are cascaded as seen in the present study suggesting that stimuli naming continues to be influenced once articulation begins, and that those factors influencing phonological aspects also influence the speed of naming. This then implies that there is in fact not a double-deficit at all.

In sum, word-effect variables on RT have shown to replicate previous findings, with word frequency being a significant predictor of RT. With regard to RD, the present study has been the first of its kind to test for word-effect variables on RD in an opaque orthography. There are three possible interpretations of these results. Firstly, the data could be confounded due to reasons mentioned previously, or those completely out of the control of the researcher. Secondly, the effects of OLD and BG mean support the idea that sub-lexical processes go beyond the syllable level as Levelt's theory proposes. Lastly, effects of OLD and word frequency suggest that phonological coding processes may continue after response onset, that is, that the phonological specification for the pronunciation of a word may not be fully prepared at response onset, providing support for a cascaded model of speech production. Which of these findings is most likely is difficult to say as this is the first report in this area. More research is needed in order to confirm these findings, however, it appears that this study might be the first of its kind in advancing the understanding of reading beyond the onset of articulation; and thus stresses the importance for reading models to also go beyond this point.

## References

- Acha, J., & Perea, M. (2008). The effects of length and transposed-letter similarity in lexical decision: Evidence with beginning, intermediate, and adult readers. *British Journal of Psychology*, 99, 245-264.
- Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, 17, 814-823.
- Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, 4, 439-461.
- Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, 55, 290-313.
- Baayen, R. H. (2008). *Analyzing linguistic data*. Cambridge, U.K.: Cambridge University Press.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.
- Balota, D. A., & Abrams, R. A. (1995). Mental chronometry: beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1289-1302.
- 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., & Yap, M. J. (2006). Attentional control and flexible lexical processing: Explorations of the magic moment of word recognition. In S. Andrews (Ed). *From inkmarks to ideas: Current issues in lexical processing*, Hove, England: Psychology Press, 229-258.
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, 39, 445-459.
- Belke, E., Brysbaert, M., Meyer, A. S., & Ghyselinck, M. (2005). Age of acquisition effects in picture naming: evidence for a lexical-semantic competition hypothesis. *Cognition*, 96, 45-54.
- Brown, G. D., & Watson, F. L. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory & Cognition*, 15, 208-216.
- Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41, 977-990.

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., & Hulme, C. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological science*, 23, 678-686.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, 63-65.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioural sciences (3rd. edition)*. Mahwah, NJ: Lawrence Erlbaum Associates.

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.

Damian, M. F. (2003). Articulatory duration in single-word speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 416-431.

Davies, R., Barbón, A., & Cuetos, F. (2013). Lexical and semantic age-of-acquisition effects on word naming in Spanish. *Memory & cognition*, 1-15.

Denckla, M. B., & Rudel, R. G. (1976). Rapid 'automatized' naming (RAN): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, 14, 471-479.

Ellis, A. W., & Lambon Ralph, M. A. (2000). Connectionist networks can provide a natural account of age of acquisition effects in adult lexical processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1103-1123.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods*, 35, 116-124.

Fredrickson, N., Frith., U & Reason, R. (1997). *Phonological Assessment Battery*. Windsor: NFER-Nelson.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, Reading Acquisition and Dyslexia: Insights from Connectionist Models. *Psychological Review*, 106, 491-528.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the Meanings of Words in Reading: Cooperative Division of Labour Between Visual and Phonological Processes. *Psychological Review*, 111, 662-720.

Hulme, C., & Snowling, M. J. (2013). Learning to Read: What We Know and What We Need to Understand Better. *Child Development Perspectives*, 7, 1-5.

Kawamoto, A. H., Goeltz, K., Agbayani, J. T., & Groel, K. (1998). Locus of semantic priming effects in speeded naming. *Psychonomic Bulletin & Review*, 5, 676-682.

Kello, C. T., Plaut, D. C., & MacWhinney, B. (2000). The task dependence of staged versus cascaded processing: An empirical and computational study of Stroop interference in speech perception. *Journal of Experimental Psychology: General*, 129, 340-360.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2013). Age-of-acquisition ratings for 30 thousand English words. *Behaviour Research Methods*, 44, 978-990.

- Levelt, W. J. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences*, 98, 13464-13471.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75.
- Manis, F. R., Doi, L. M., & Bhadha, B. (2000). Naming speed, phonological awareness, and orthographic knowledge in second graders. *Journal of learning disabilities*, 33, 325-333.
- Meyer, A. S., & Damian, M. F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory & Cognition*, 35, 494-503.
- Monaghan, J., & Ellis, A. W. (2002). Age of acquisition and the completeness of phonological representations. *Reading and Writing*, 15, 759-788.
- Morrison, C. M., & Ellis, A. W. (1995). The roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 21, 116-133.
- New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Re-examining the word length effect in visual word recognition: New evidence from the English lexicon project. *Psychonomic Bulletin and Review*, 13, 45-52.
- Pineiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-plus (Statistics and Computing)*. New York: Springer.
- Plaut, D. C., McClelland, J. L., Seidenberg, M., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103, 56-115.
- Protopapas, A. (2007). Check Vocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, 39, 859-862.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.
- Seidenberg, M. S., & Plaut, D. C. (1998). Evaluating word-reading models at the item level: Matching the grain of theory and data. *Psychological Science*, 234-237.
- Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, 411-416.
- Steyvers, M., & Tenenbaum, J. B. (2005). The large scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cognitive Science*, 29, 41-78.
- Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1140–1154.
- Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., et al. (1999). Preventing reading failure in your children with phonological



processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology*, 91, 579–593.

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 50A, 439–456.

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of educational psychology*, 91, 415-438.

Wolf, M., Bowers, P. G., & Biddle, K. (2000). Naming-speed processes, timing, and reading. A conceptual review. *Journal of Learning disabilities*, 33, 387-407.

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, 15, 971-979.