Journal of Experimental Psychology:
Learning, Memory, and Cognition
2016, Vol. 42, No. 4, 542–565

© 2015 American Psychological Association
0278-7393/16/$12.00 http://dx.doi.org/10.1037/xlm0000191

Short-Term and Long-Term Effects on Visual Word Recognition

Athanassios Protopapas and Efthymia C. Kapnoula
University of Athens

Effects of lexical and sublexical variables on visual word recognition are often treated as homogeneous across participants and stable over time. In this study, we examine the modulation of frequency, length, syllable and bigram frequency, orthographic neighborhood, and graphophonemic consistency effects by (a) individual differences, and (b) item repetition. A group of 129 participants performed lexical decision and naming, in counterbalanced order, using a set of 150 Greek words in which these variables were decorrelated. Frequency, length, and syllable frequency effects were reduced by a preceding task. Length effects were inversely related to years of education. Neighborhood effects depended on the metric used. There were no significant effects or interactions of bigram frequency or consistency. The results suggest that exposure to a word causes transient effects that may cumulatively develop into permanent individual differences. Models of word recognition must incorporate item-specific learning to account for these findings.

Keywords: visual word recognition, Greek, frequency, length, syllable frequency

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

How is the visual word recognition system affected by the processing of a word? In this study, we approach this question by considering the short-term and long-term effects of visual word recognition on the processing system itself. Models of visual word recognition posit mechanisms and representations involved in the processing of visual orthographic stimuli. The implications of these hypotheses are typically studied using naming and lexical decision tasks. A productive line of research concerns the effects of lexical and sublexical variables—such as frequency, length, neighborhood, bigram and syllable frequency, and more—on distributions of response time (RT; see reviews in Balota, Yap, & Cortese, 2006, and Balota, Yap, Hutchison, & Cortese, 2012).

An early approach based on groups of stimuli differing in parameters of interest has been criticized for selection of atypical items and restricted parameter ranges (Balota et al., 2012). More recent approaches are based on multivariate analyses of several variables examined simultaneously (Yap & Balota, 2009) and on very large databases consisting of thousands of words responded to by hundreds of participants (e.g., Balota et al., 2007; Ferrand et al., 2010; Keuleers, Diependaele, & Brysbaert, 2010; Keuleers, Lacey, Rastle, & Brysbaert, 2012). The observed effects of the studied parameters are used to evaluate models of visual word recognition and reading aloud in an attempt to address issues of theoretical interest (e.g., Adelman & Brown, 2008a, 2008b; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Mulatti, Reynolds, & Besner, 2006; Perry, Ziegler, & Zorzi, 2007, 2010; Reynolds & Besner, 2002, 2004). By and large, these investigations have examined stable undifferentiated effects of the lexical and sublexical variables as if they were fixed in time and homogeneous across participants.

Long-Term Effects

Individual differences in the magnitude of these effects have recently attracted some attention as data from large-scale studies have become available, and as more sophisticated statistical methods (Baayen, 2008, 2013; Baayen, Davidson, & Bates, 2008) permit distinctions between sources of variance attributable to persons, items, and their properties (Adelman, Sabatos-DeVito, Marquis, & Estes, 2014; Kuperman, Drieghe, Keuleers, & Brysbaert, 2013; Kuperman & Van Dyke, 2011, 2013; Yap, Balota, Sibley, & Ratcliff, 2012). These advances parallel studies of priming in which the realization of individual differences is also becoming theoretically attractive (e.g., Andrews & Hersch, 2010; Andrews & Lo, 2012). Rising to the occasion, individual-level computational modeling has emerged as a promising direction to investigate systematic patterns of relations among variables in relation to specific theoretical architectures (Adelman et al., 2014; Ziegler, 2011; Ziegler et al., 2008).

The effect of word frequency is the one most studied. It has been found to be greater for participants with overall slower responses (Schilling, Rayner, & Chumbley, 1998) or with relatively lower reading skill (Ashby, Rayner, & Clifton, 2005; Kuperman & Van Dyke, 2011; but depending on the frequency metric, cf. Kuperman & Van Dyke, 2013), fewer years of formal education (Tainturier, Tremblay, & Lecours, 1992), lower print exposure (Chateau & Jared, 2000; but only when pseudohomophones were used in a lexical-decision task, cf. Lewellen, Goldinger, Pisoni, & Greene, 1993; Sears, Siakaluk, Chow, & Buchanan, 2008), lower vocabu-

This article was published Online First October 5, 2015.

Athanassios Protopapas, Department of Philosophy and History of Science, University of Athens; Efthymia C. Kapnoula, Graduate Program in Basic and Applied Cognitive Science, University of Athens.

Efthymia C. Kapnoula is now at the Department of Psychological and Brain Sciences, University of Iowa.

We are grateful to Victor Kuperman for comments on a previous version of the manuscript, and to Harald Baayen, Pierre Courrieu, Emmanuel Keuleers, Michael Cortese, and anonymous reviewers for constructive criticism and helpful suggestions.

Correspondence concerning this article should be addressed to Athanassios Protopapas, MITHE, Ano Ilissia University Campus, GR-157 71 Zografos, Greece. E-mail: aprotopapas@phs.uoa.gr

lary (Yap et al., 2012), or lower language proficiency, for monolinguals and bilinguals alike (Diependaele, Lemhöfer, & Brysbaert, 2013). Thus, it seems that frequency effects may diminish with higher facility in verbal skills, including oral language and reading. On the other hand, older participants have been found to exhibit greater frequency effects (for lower frequency words) than younger participants in a lexical decision task, a difference attributed to increased vocabulary associated with protracted language experience (Ramscar, Hendrix, Love, & Baayen, 2013; Ramscar, Hendrix, Shaoul, Milin, & Baayen, 2014). Thus, overall, existing findings are far from uniform, hinging on metrics and stimulus properties, and remain equivocal as to the source of the differences, not clearly distinguishing between expertise (related, e.g., to reading exposure) and verbal ability (as indicated, e.g., by vocabulary).

Increased effects of word length have also been reported for participants with lower reading skill (Kuperman & Van Dyke, 2011) or lower vocabulary (Yap et al., 2012). Yap et al. (2012) also found increased neighborhood size effects associated with lower vocabulary scores and interpreted their overall pattern of results as "consistent with skilled readers being more reliant on relatively automatic lexical processing mechanisms, and hence showing less influence of word characteristics" (p. 69). These effects have not been compared across variables, and it remains unclear whether they may be attributed to a common source or to distinct developmental processes.

To the extent that individual differences can be conceptualized as cumulative long-term effects resulting from reading experience (e.g., "changes to the connection weights between lexical–lexical and lexical–semantic modules"; Blais, O'Malley, & Besner, 2011), it is instructive to investigate them in more detail, aiming to delineate constraints and directions for modeling the development of the reading apparatus. The developmental view has been central in the connectionist tradition of reading models (Harm, McCandliss, & Seidenberg, 2003; Harm & Seidenberg, 1999, 2004; Plaut, 2006) but largely neglected in the dual-route approach (but see Ziegler, Perry, & Zorzi, 2014, for a recent attempt to remedy the situation). The study of individual differences attributable to developmental history can provide additional testing grounds and suggest modifications to the models toward capturing cognitive function more realistically. Moreover, as most existing studies have focused on the effect of word frequency, it would be instructive to examine a wider range of effects within a common analytic framework.

Short-Term Effects

In addition to stable long-term individual differences, short-term effects arising from experimental manipulations, such as repetition, can also be instructive. After all, how could long-term individual differences arise if not from repeated short-term effects of exposure, accumulated over extended periods? For example, repeated reading is long known to increase reading fluency with each repetition (e.g., Lemoine, Levy, & Hutchinson, 1993; Levy, Nicholls, & Kohen, 1993; cf. "local frequency" effects in Baayen & Milin, 2010), consistent with episodic transfer between repeated encounters with specific words (Levy, Di Persio, & Hollingshead, 1992). Well-known short-term effects on visual word recognition are generally referred to as "priming," hypothesizing that prior activation of specific lexical items affects subsequent processing of those same items or other items related to them (e.g., orthographic neighbors). This theoretical approach has been very fruitful in highlighting types of relations among items and potential mechanisms underlying various effects on lexical processing often understood as facilitation or inhibition. However, there has been little opportunity to investigate whether and how the effects of recent—not necessarily immediate—involvement with specific items may be modulated by stable properties of the items other than their relations to each other.

Repeated exposure has been found to diminish effects of frequency (with isolated word naming and lexical decision, see Blais et al., 2011, Colombo, Pasini, & Balota, 2006, and Visser & Besner, 2001; but not with text reading, cf. Raney & Rayner, 1995), stimulus quality (Blais & Besner, 2007), and regularity (in lexical decision, but not in naming; Katz et al., 2005). Adelman, Marquis, Sabatos-De Vito, and Estes (2013) and Keuleers et al. (2010) found little effect of repeated practice—in isolated word reading and in lexical decision, respectively—on the effects of individual variables, namely, a trend toward weakening, consistent with the previous reports. However, both of these studies focused on overall trends over sessions, not analyzing their data in terms of repetition of individual items.[1] Finally, Mulatti, Peressotti, Job, Saunders, and Coltheart (2012) found increasing response times with repeated readings of orthographic neighbors, which they interpreted as cumulative effects of lexical competition.

There is, therefore, a small but growing literature on the effects of repeated processing of the same stimuli, which, at the moment, has examined few and scattered effects, primarily related to word frequency. The potential importance of these studies for models of word recognition is tremendous, because they can help disentangle loci of effects in relation to modeled representations, processes, and even entire modules (cf. Blais et al., 2011). Not only must models demonstrate effects of "delayed priming" owing to recent processing of specific items—they must also be able to produce the observed facilitation or inhibition differentially across items, in accordance with their individual properties (frequency, length, etc.). For example, if frequency effects are reduced by repetition, then models of visual word recognition must exhibit more delayed priming for low-frequency words than for high-frequency words. In conjunction with constraints arising from long-term effects of cumulative experience, the analysis of short-term effects of repetition can be useful to modelers, especially if it concerns a wide range of variables examined under a systematic approach.

Notably, the existing literature outlined above, on both long-term and short-term effects, has been dominated by studies in the English language, a well-known outlier in terms of orthographic consistency (Share, 2008), with implications for the rate, processing strategies, and underlying representations of reading development (Seymour, Aro, & Erskine, 2003; Ziegler & Goswami, 2005, 2006), and, possibly, for the cognitive

[1] Keuleers et al. (2010) did analyze an identical repeated block, with inconclusive results, but this cannot be said to address short-term effects because the two repetitions were spaced several weeks apart (as noted in Diependaele, Brysbaert, & Neri, 2012).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

processes of adult skilled visual word recognition. Therefore, systematic studies in other languages can also be useful in confirming or challenging the cross-linguistic generality of existing findings and, thereby, in highlighting potentially relevant factors underlying short-term and long-term effects within and across languages.

The Current Approach

In the present study we exploit the dual-task experimental procedure, using both naming and lexical decision, to investigate the extent to which recent exposure to the same set of words in a related but distinct task differentially affects word processing in the subsequent task. Our short-term effect variable, then, is task order: Half of our participants completed lexical decision before naming, and vice versa for the other half. In other words, each task was administered first to one half of the participants and second to the other half. Despite some attempt to minimize carryover effects with a distractor interim task, the words used were arguably primed from the first task. One goal of this report is to identify whether such priming is not only significant but also systematically related to the variables under study.

The second main goal of the study concerns long-term effects. Specifically, we were interested in whether differences in age, education, or reading skill (as quantified by overall naming and lexical decision performance) are related to the effects of lexical and sublexical variables on naming and lexical decision performance. Our participants were all skilled adult readers (high school graduates); therefore, the study falls squarely within the purview of existing word recognition models, rather than in the potentially distinct domain of reading development. That is, we are interested in studying how relatively stable expert reading systems are affected by the results of their own processing over the short term and the long term, rather than how properties of immature reading systems develop on the basis of gradually accumulating experience.

Using our diverse set of variables, we analyzed the short-term and long-term effects in a unified and systematic approach. We work in the Greek language, which has a well-understood, relatively transparent orthographic system (Protopapas & Vlahou, 2009), aiming to extend the range of investigation toward more inclusive cross-linguistic validity. The existence and magnitude of the basic effects of lexical and sublexical variables on naming and lexical decision RT are not of primary concern in the present report, as they have been previously established (Protopapas & Kapnoula, 2013). Following up on those findings, the specific focus of this study concerns the *differences* in the effects of certain well-studied variables as a function of (a) recent, and (b) presumed overall, exposure to and processing of specific words.

Method

One persistent difficulty in analyzing the effects of lexical and sublexical variables on naming and lexical decision RTs originates in the significant intercorrelations among many variables, which make it difficult to disentangle effects specifically attributable to each variable. Between-groups matching, multiple mixed-effects modeling (Adelman et al., 2013, 2014), analysis of residuals (Kuperman & Van Dyke, 2011), hierarchical multiple regression (Yap & Balota, 2009), principal components approaches (Yap et al., 2012), and combinations of the above (Baayen, Feldman, & Schreuder, 2006) have been used to address this problem to some extent (with occasional problems of interpretation; see, e.g., Wurm & Fisicaro, 2014, on residualizing). Compounding the intercorrelation problem, interactions among variables and potential nonlinearities in the individual effects preclude simple linear regression-based approaches from effectively sorting out specific effects, because it becomes impossible to distinguish nonadditivity due to interactions from that due to nonlinearities when the variables are correlated. Even if the shape of an effect is approximately known, "removing" it statistically via an appropriate regressor has the undesirable side effect of also partially removing effects of other variables that are correlated with it. Conversely, if the shape of an effect is not precisely known, arbitrarily removing a linear effect will leave behind higher order residuals that can then be falsely attributed to other correlated variables. Thus, the intercorrelations among variables constitute a potential impediment to our understanding of their individual effects and, hence, to their effective modeling.

To address these issues from a different perspective, we have focused on stimulus selection, aiming to reduce, as much as possible, the intercorrelations among variables of interest. That is, we have selected words—and constructed pseudowords—to form a set within which the correlations of predictor variables are very low and not significant. This option is not a panacea, as it introduces its own set of problems (including the risk of selecting atypical stimuli), so we are not claiming that it is always superior and should replace all other methods. However, it is a methodological approach complementary to the statistical approaches that are more usually found in the literature. Using this method, we can begin to distinguish among effects of interest, directly testing the linearity and additivity assumptions, and to determine the most suitable metric and transformation scales for further investigation using other, more established approaches. In Appendix A, we present simulations demonstrating that decorrelating variables permits more accurate independent estimation of individual effects and reduces artifactual interactions arising from unmodeled nonlinearities.

We have chosen to examine a diverse set of lexical and sublexical variables, including frequency, length, syllable frequency, neighborhood, bigram frequency, and graphophonemic consistency. As previously documented (Protopapas & Kapnoula, 2013), the effects of these variables can be examined individually in separate analyses, greatly simplifying model estimation and testing of particular coefficients. Our stimuli span a wide range on each studied variable and, despite the constraints on their selection, may be more representative of the participants' vocabulary than other artificially selected sets (e.g., monosyllables, often used in English). Therefore, this set of stimuli is uniquely suitable for the study of the effects of these variables, as they may be modulated by short-term manipulations and by long-term individual differences.

Participants

The sample included 97 women and 35 men, native speakers of Greek, 18 to 36 years old (M = 23.3, SD = 4.7). Most were undergraduate or graduate students (12 to 21 years of education; M = 15.4, SD = 2.1). Fourteen were left-handed. The number of

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

participants was based on the analysis of Rey, Courrieu, Schmidt-Weigand, and Jacobs (2009), aiming to maximize the reliability of item variance.

Materials

A set of 150 words were selected from the ILSP PsychoLinguistic Resource (IPLR) word list (speech.ilsp.gr/iplr; Protopapas, Tzakosta, Chalamandaris, & Tsiakoulis, 2012), two to five syllables long, spanning a wide range over several target variables. Table 1 lists their descriptive statistics compared with the entire corpus. Orthographic and phonological syllable and bigram frequency refer to the mean logarithmic token frequency of (position-independent) syllables or symbol pairs (letters or phonemes), respectively, in occurrences per million tokens. Orthographic and phonological neighborhood counts refer to Coltheart's N, that is, the number of words with the same length that differ by only one letter or phoneme, respectively (Coltheart, Davelaar, Jonasson, & Besner, 1977). Graphophonemic consistency was computed as the logarithmic mean of nondirectional token "sonograph" probabilities, that is, ratios of specific grapheme-phoneme mappings over the total number of grapheme–phoneme tokens (Spencer, 2009). In an iterative process, items were selected and a nonparametric index of association (Spearman's ρ) among all variables was calculated; the process terminated when groups of qualitatively distinct variables were not significantly correlated. The final intercorrelations among variables in the selected items are shown in Table 2. The distributions of stimulus variables compared with corpus type and token distributions are shown in Appendix B.

A set of 150 pseudowords were constructed to resemble the words in basic phonological and orthographic structure and in letter and phoneme distribution. The pseudowords were indistinguishable from the words in the target variables, as verified by the Kolmogorov–Smirnov test for equality of distributions. The results of these tests are also listed in Table 1, and intercorrelations are shown in Table 2. The full list of stimuli is presented in Appendix C.

Pseudoword neighborhoods were excluded from matching and intercorrelation requirements to avoid undue activation of specific lexemes by pseudoword stimuli (especially because neighbors of long pseudowords might be inflectional variants of a single base form). Thus, pseudowords were constructed with as few neighbors as possible. As a result, no claims can be made about effects on pseudoword processing that are correlated with neighborhood size. This is not an issue in the present study, as we focus on the analysis of responses to words only. However, the fact that pseudowords have few or no neighbors may have implications for the interpretation of effects on words in the lexical-decision task.

Procedure

A naming and a lexical-decision task were implemented in DMDX (Forster & Forster, 2003). In both tasks, each item was presented in black Arial 36-pt white font on a white background at the center of a 15.4-in. laptop screen for 2,000 ms. A few practice and warm-up items preceded the experimental stimuli. A short break was offered every 75 stimuli.

For lexical decision, participants responded by pressing the left and right control keys. Words and pseudowords were intermixed randomly. The "word" response was set to the participant's preferred or nonpreferred hand, approximately counterbalanced across participants. For naming, words and pseudowords were presented in separate blocks, in counterbalanced order between participants. Responses were recorded via a headset and onset times were subsequently verified using CheckVocal (Protopapas, 2007). The order of naming and lexical decision tasks was counterbalanced. In both tasks, items were presented in a different random order for each participant. A distractor task (digit span) was administered between the two tasks to reduce carryover effects.

Data Analysis

Raw response times (in milliseconds, for correct responses only) were inverted (transformed to −1,000.0/RT), rather than logarithmically transformed, as this resulted in better approximation to the normal distribution both for the raw data and for the model residuals (cf. Baayen & Milin, 2010). The transformed RTs were

Table 1
Descriptive Statistics for Words, Pseudowords, and Corpus Types and Tokens, for Each Variable, and Results From the K-S Test Comparing Word and Pseudoword Distributions

	Words			Pseudowords		K-S test		Corpus types		Corpus tokens	
	M	*SD*	Range	*M*	*SD*	*D*	*p*	*M*	*SD*	*M*	*SD*
Log frequency[a]	.88	1.89	−3.38 to 4.22					−1.49	1.81	6.22	3.27
Number of letters[a]	7.24	1.93	4 to 10	7.31	1.84	.060	.950	10.07	2.71	5.43	3.17
Number of phonemes	7.09	1.93	4 to 11	7.00	1.82	.053	.983	9.45	2.61	5.02	3.03
Number of syllables	3.11	.95	2 to 5	3.10	.92	.020	1.000	4.35	1.29	2.38	1.45
Orth. neighbors[a]	2.17	1.50	0 to 7	.39	1.21	.700	<.001	1.38	1.58	5.88	4.83
Phon. neighbors	3.25	1.88	0 to 10	1.31	3.31	.633	<.001	2.36	3.27	10.72	8.42
Orth. bigram frequency[a]	.76	.26	.22 to 1.31	.79	.32	.093	.531	1.02	.42	1.91	1.46
Phon. bigram frequency	1.03	.37	.23 to 2.03	.98	.44	.140	.106	1.23	.59	2.10	1.63
Orth. syllable frequency	7.86	5.70	.85 to 20.1	6.36	4.73	.153	.059	8.93	6.08	11.64	8.23
Phon. syllable frequency[a]	11.07	5.78	.92 to 21.7	9.82	5.40	.147	.079	12.61	6.66	15.89	9.40
G-P consistency[a]	32.67	8.03	15.2 to 49.5	32.75	7.68	.053	.983	32.13	9.41	38.84	15.66

Note. K-S = Kolmogorov–Smirnov; Orth. = orthographic; Phon. = phonological; G-P = graphophonemic.
[a] Variable subsequently selected for the "experimental" set.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 2
Nonparametric Correlation Coefficients (Spearman's ρ) Between Variables for Words and Pseudowords

	2	3	4	5	6	7	8	9	10	11
1. Log frequency[a]	−.049	−.084	−.084	−.002	.069	.104	.091	−.081	−.024	.002
2. Number of letters[a]		.965*	.860*	−.007	.039	.008	.034	−.080	−.045	−.106
3. Number of phonemes	.947*		.872*	.057	.023	−.067	.089	−.096	−.100	−.057
4. Number of syllables	.870*	.904*		.044	.086	−.096	.042	.017	.082	−.050
5. Orth. neighbors[a]	−.504*	−.505*	−.486*		.640*	.071	.084	−.043	−.081	−.018
6. Phon. neighbors	−.612*	−.659*	−.649*	.685*		.133	.095	−.046	−.017	−.052
7. Orth. bigram frequency[a]	.083	−.060	−.056	−.037	.106		.326*	.060	.086	.056
8. Phon. bigram frequency	−.055	.039	.000	.100	.177*	.214*		−.253*	−.068	.110
9. Orth. syllable frequency	−.065	−.010	.124	.096	.019	.092	.017		.786*	.064
10. Phon. syllable frequency[a]	−.077	−.081	.106	.065	.056	.065	.101	.697*		−.093
11. G-P consistency[a]	.003	.122	.036	−.032	−.121	.046	.122	.044	−.123	

Note. Coefficients for words are above the diagonal; those for pseudowords are below the diagonal. Orth. = orthographic; Phon. = phonological; G-P = graphophonemic.
[a] Variable subsequently selected for the "experimental" set.
* $p < .05$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

analyzed in R version 3.0.2 (R Core Team, 2013) using linear mixed-effects models with crossed random effects for participants and items (Baayen, 2008; Baayen et al., 2008). Models were estimated with full maximum likelihood using package lme4 (Bates, Maechler, Bolker, & Walker, 2013). For significance testing, *p* values were calculated with Satterthwaite's approximation for the fixed effects, using package lmerTest (Kuznetsova, Brockhoff, & Christensen, 2013), and with log-likelihood ratio tests for the random effects. All analyses involving naming data included fixed effects of first and second phoneme class (six categories: fricative, liquid, nasal, voiced stop, unvoiced stop, and vowel). All independent variables were centered. All nominal variables were deviation coded (see Appendix D for coding polarity).

Results

Data from two participants who failed to complete lexical decision, and from one participant with missing demographic information, were removed, leaving full data sets in both tasks for 129 participants to be further analyzed. Trials with no response within the 2-s time-out period or with an RT less than 200 ms were marked as incorrect. Mean accuracy was .97; accuracy per item ranged between .67 and 1.00 (SD = .06); accuracy per participant ranged between .89 and 1.00 (SD = .02).

A series of omnibus models were estimated first, including data from both tasks. Fixed effects included linear effects and interactions among experimental (lexical and sublexical) variables, as well as task variables (task type, trial order, task order, preceding RT) and participant variables (age, sex, education, handedness). Random structures included random slopes per participant for the linear effects of trial number and preceding RT, in addition to random intercepts for participants and items (words). Random slopes of task were also included for both participants and items, fully interacting with the other random effects. Models were gradually trimmed to remove nonsignificant fixed effects. It was confirmed that (a) age, sex, handedness, and phonological neighborhood had no significant effects or interactions; (b) item length effects were completely represented by the number of letters; and (c) syllable and bigram effects were limited to their phonological and orthographic versions, respectively.

The resulting model was subjected to criticism focusing on the residuals (Baayen & Milin, 2010). Data points (n = 185; 0.5% of the data) with an absolute value of standardized residual exceeding 3.29 (corresponding to $p < .001$) were removed from the data set and the model was reestimated, resulting in greatly improved fit (deviance = −1,804, down from 986; BIC = −933, from 1,858) with minor or negligible effects on coefficient estimates. All subsequent analyses employ the trimmed data set.

To confirm the selection of a variable to play the role of a long-term individual differences proxy, we fit a baseline model with just a fixed intercept and random intercepts for participants and items. Subsequent models including a fixed effect of age, education, or both were compared with the baseline. Age accounted for 4.1% of participant variance (a statistically significant effect, $\beta = -8.83 \times 10^{-3}$, $t = -2.32$, $p = .022$) when entered alone, and education for 7.7% ($\beta = -2.76 \times 10^{-2}$, $t = -3.26$, $p = .001$). However, the effect of age was completely subsumed by education, as the participant variance accounted for by both was 7.9%, consistent with the effect of age being suppressed and nonsignificant in the presence of education ($\beta = 3.21 \times 10^{-3}$, $t = 0.50$, $p = .618$). Therefore, the data indicated that education—but not age—was a relevant between-participants variable, accounting for sufficient variance in this adult population, justifying its selection for further analyses. Despite the limited variability one might expect in the number of years of education among a population of school graduates, our participants spanned a rather wide range (12 to 21 years), permitting robust effects to emerge.

Interactions of task type (naming vs. lexical decision) with the effects analyzed below were not removed in order to permit comparisons across tasks. A complete model including all remaining factors is listed in Appendix D. In this model, there were strong parametric effects, primarily of frequency and length, and, to a smaller extent, of other variables. Task type, task order, and education interacted significantly with some of these parametric effects, indicating that recent and long-term experience with specific words affects the way they are processed, and that the magnitude of this influence differs between the two tasks. The significant interactions of task necessitated breaking down further analyses, separating lexical decision from naming.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Although the pseudowords were constructed based on real words and were matched to the words in bigram and syllable frequency, the fact that they were not matched in neighborhood size leaves open the possibility that they may have been sufficiently dissimilar to the words to bias the lexical decision task and allow it to be performed without reliance on lexical activation. To alleviate this concern, we used the LD1NN algorithm to quantify the bias inherent in our stimulus set (Keuleers & Brysbaert, 2011), applied to the specific stimulus sequence delivered to each of the 129 participants using R package vwr (Keuleers, 2013). No significant bias was detected (mean odds = 0.82, range = 0.71 to 0.99; mean $z = -1.12$, range = −1.93 to −0.08). In particular, cumulative average bias was negligible for words, whereas pseudowords were somewhat biased toward words, confirming their word-like construction. Including the word probabilities estimated by the LD1NN algorithm in the RT analysis model for the lexical decision task produced no significant effect ($\beta = -3.72 \times 10^{-3}$, $t = -1.43$, $p = .153$) and no interaction with trial order ($\beta = 3.69 \times 10^{-3}$, $t = 1.43$, $p = .154$). Therefore, this factor was not included in subsequent analyses.

Six variables were selected to represent lexical and sublexical effects: log frequency (frequency [F]), number of letters (length [L]), mean phonological syllable frequency (syllable frequency [S]), orthographic neighborhood size (neighborhood [N]), orthographic bigram frequency (bigram frequency [B]), and graphophonemic consistency, as defined above (consistency [C]). These are henceforth termed "experimental" variables. As noted above, these variables were not significantly correlated with one another, and therefore their effects can be investigated in isolation. The modulation of the effects of each of these variables in each task (T; lexical decision vs. naming) by the short-term proxy task order (TO; first vs. second) and by the long-term proxy education (E; scale, centered) was examined in a series of models including (a) a set of generic fixed effects, specifically, trial order (Ord) and RT to the preceding item (invRT1; to account for sequential effects, cf. Baayen & Milin, 2010; Taylor & Lupker, 2001); (b) the fixed linear effect of one experimental variable interacting with task order, education, and their interaction; and (c) a random structure including intercepts for participants (sID) and items (iID), as well as a random slope for the experimental variable for participants, correlated with the participant random intercept, and an uncorrelated random slope for trial order. For example, the frequency model (in R notation) was specified as

$$\text{invRT} \sim \text{TaskOrd} * \text{Education} * \text{logfreq} + \text{invRT1} + \text{Ord} \\ + (1 + \text{logfreq}|\text{sID}) + (0 + \text{Ord}|\text{sID}) + (1|\text{iID}).$$

Table 3 shows the results of fitting these models for each task and experimental variable. To help interpret these results, Figures 1 and 2 display the corresponding interaction plots, created with the effects package (Fox, 2003). In each panel, the effect of an experimental variable is evaluated at the two levels of task order and at endpoint values of education ("Low" = 12 years; "High" = 21 years). The vertical range spanned by the predicted variable in each panel is displayed numerically in order to facilitate comparisons.

Clear effects emerged for frequency and length, in both naming and lexical decision, as expected, and in syllable frequency, only in lexical decision. In addition, there were significant interactions of these variables with education and task order, but no significant triple interaction.

Specifically, the effect of frequency was negative, indicating faster responses to more frequent words. This was modulated by task order such that it was reduced by a preceding task. The modulation of the frequency effect by task order did not differ between tasks (in the full model, interaction T × TO × F, $\beta = -2.92 \times 10^{-3}$, $t = -1.12$, $p = .262$).

The effect of length was positive, indicating slower responses to longer words. This was more pronounced in participants with fewer years of education. It was also modulated by task order such that it was reduced by a preceding task. The modulation of the length effect by education and task order did not differ between tasks (in the full model, T × TO × L, $\beta = 1.27 \times 10^{-3}$, $t = 0.50$, $p = .617$; T × E × L, $\beta = 7.74 \times 10^{-4}$, $t = 1.27$, $p = .203$). The combination of the two factors in lexical decision led to the notable situation, seen in the top right panel of Figure 1, of a negligible length effect estimate for the most highly educated participants who had previously completed the naming task.

The effect of syllable frequency was also positive, indicating slower responses to words with more frequent syllables. Like the frequency and length effect, the syllable frequency effect was reduced by a preceding task; in naming this rendered the main effect nonsignificant. The modulation of the syllable frequency effect by task order did not differ between tasks ($\beta = 4.19 \times 10^{-4}$, $t = 0.49$, $p = .622$). In naming, there was also an interaction of syllable frequency with education, consistent with a smaller effect for the participants with more years of education. The difference between tasks in the modulation of the syllable frequency effect by education approached significance ($\beta = -3.95 \times 10^{-4}$, $t = -1.94$, $p = .052$).

The effects of the other three variables were less reliable. In particular, there was no effect of neighborhood, bigram frequency, or consistency in naming, and no interaction with education or task order. In lexical decision, there was a marginally significant main effect of neighborhood, in the direction of faster responses to words with more neighbors. In addition, there was a significant interaction of neighborhood with task order ($\beta = -5.34 \times 10^{-3}$, $t = -3.91$, $p < .001$), indicating that the effect of neighborhood was only reliable when the lexical-decision task was administered first, and was diminished by a preceding naming task. Accordingly, the triple interaction between task, task order, and neighborhood was significant ($\beta = -9.09 \times 10^{-3}$, $t = -2.80$, $p = .005$). There was no significant main effect of bigram frequency in lexical decision, but only an interaction with task order ($\beta = 1.69 \times 10^{-2}$, $t = -2.18$, $p = .029$), consistent with a small bigram frequency effect only when preceded by the naming task. Finally, there was no significant effect or interaction of consistency in lexical decision. There was no significant interaction of the modulation of the effects of bigram frequency or consistency by task order or education with task (triple interactions in the full model; all $p > .1$).

Turning to the correlations between participant random intercepts and random slopes of the experimental variables, they were significant for frequency, in both tasks, for length, in lexical decision (and marginally in naming), and for syllable frequency, in naming only. In each of these cases, the correla-

Table 3
Main Effects and Interactions With Education and Task Order on Naming and Lexical Decision RT Involving the Six Experimental variables

	Main effect			Interactions: × Task order			Interactions: × Education			Interactions: × Task order × education (triple)			Random slope correlation		
	β	*t*	*p*	β	*t*	*p*	β	*t*	*p*	β	*t*	*p*	*r*	χ^2	*p*
Full models															
Lexical decision															
Frequency	−42	−7.7	.000	−4	−2.9	.005	−1	−1.0	.341	−.3	−.5	.648	.375	7.15	.008
Length	20	3.2	.002	5	3.4	.001	−3	−3.4	.001	.3	.5	.642	.248	4.30	.038
Syllable frequency	6	3.1	.002	1	3.2	.002	.1	.7	.501	.1	.4	.716	−.079	.14	.709[a]
Neighborhood	−14	−1.7	.090	−5	−3.9	.000	−.5	−.8	.452	1	1.9	.056	.575	1.66	.197[a]
Bigram frequency	−14	−.3	.759	17	2.2	.029	−5	−1.3	.144	−.1	<.1	.988	1.000	—	—[a]
Consistency	.8	.6	.583	.4	1.5	.136	.04	.3	.776	.06	.4	.653	−1.000	—	—[a]
Naming															
Frequency	−26	−6.1	.000	−3	−3.2	.002	−.04	−.1	.927	.1	.3	.772	.323	4.82	.028
Length	27	6.5	.000	4	3.0	.003	−2	−2.3	.021	−.1	−.1	.891	.190	3.48	.062
Syllable frequency	2	.8	.430	1	3.2	.002	−.3	−1.9	.061	<.01	<.1	.973	.520	12.39	<.001
Neighborhood	−3	−.5	.625	−1	−1.1	.290	−.2	−.4	.716	.1	.1	.892	−1.000	—	—[a]
Bigram frequency	−21	−.6	.546	6	1.2	.248	−3	−1.0	.305	2	.6	.557	1.000	—	—[a]
Consistency	−.2	−.1	.886	.2	1.3	.196	.01	.1	.883	.01	.2	.876	−1.000	—	—[a]
Models excluding nonsignificant effects															
Lexical decision															
Frequency	−41	−7.7	.000	−4	−2.8	.006	−1	−1.1	.273				.384	7.15	.008
Length	19	3.2	.002	5	2.9	.004	−2	−2.9	.004				.257	4.30	.038
Syllable frequency	6	3.2	.002	1	3.5	.000	.1	.7	.502						
Neighborhood	−14	−1.7	.086	−6	−4.1	.000	−1	−.9	.383						
Bigram frequency	−15	−.3	.733	16	2.0	.049	−5	−1.3	.181						
Consistency	.8	.5	.591	.3	1.0	.342	.04	.3	.741						
Naming															
Frequency	−26	−6.1	.000	−3	−3.1	.002	.0	.0	.969				.297	4.83	.028
Length	27	6.5	.000	4	2.9	.005	−2	−2.4	.018						
Syllable frequency	2	.8	.445	1	2.7	.009	−.4	−2.5	.013				.443	12.39	<.001
Neighborhood	−3	−.6	.578	−1	−1.0	.341	−.2	−.5	.643						
Bigram frequency	−21	−.6	.554	6	1.1	.251	−2	−.9	.354						
Consistency	−.2	−.1	.891	.2	1.2	.225	.04	.4	.688						

Note. Results for full models in the top half of the table; results for models excluding the triple interaction and the random effects slope and correlation, when nonsignificant ($p > .05$ in 1-*df* χ^2 comparison), in the bottom half. Frequency = log tokens; length = number of letters; phonological syllable frequency; orthographic neighborhood; orthographic bigram frequency; graphophonemic consistency; all β values multiplied by 1,000. Task order effects are positive from first to second.
[a] Random slope not significant.

tion coefficient was positive, indicating individual differences in the effect of the experimental variable over and above the effects of education and task order. For length and syllable frequency, the effects of which were also positive, the sign of the correlation indicates that the effect of the experimental variable was larger for slower participants (i.e., with greater mean RTs). In contrast, for frequency, which had a negative effect, the sign of the correlation indicates that the effect of the experimental variable was smaller for slower participants. Some correlations were estimated at ±1.00, indicating overparameterization (associated with nonsignificant random slopes of the experimental variable), and are not considered further. Table 3 (lower portion) lists the results for models excluding nonsignificant triple interactions and random effects.

Discussion

Our analyses have revealed that visual word recognition is affected by the reader's recent and accumulated exposure to lexical items. Specifically, both short-term and long-term effects were revealed for length. Short-term effects also emerged for frequency and syllable frequency. These modulations did not differ between naming and lexical decision. Evidence for short-term effects was obtained for neighborhood and bigram frequency, in lexical decision only, subject to reservations stemming from the large number of comparisons performed without statistical correction to control familywise error rate. There were no effects involving graphophonemic consistency in either task.

To our knowledge, this is the first simultaneous systematic investigation of short-term and long-term effects on a diverse set of variables often examined in word recognition research. Because of the large sample and well-controlled stimulus set, the results can be constructively interpreted in the context of computational models of word recognition, providing novel directions for future improvements. In the remainder of this section, we discuss our findings and their broader implications in the context of the current literature.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

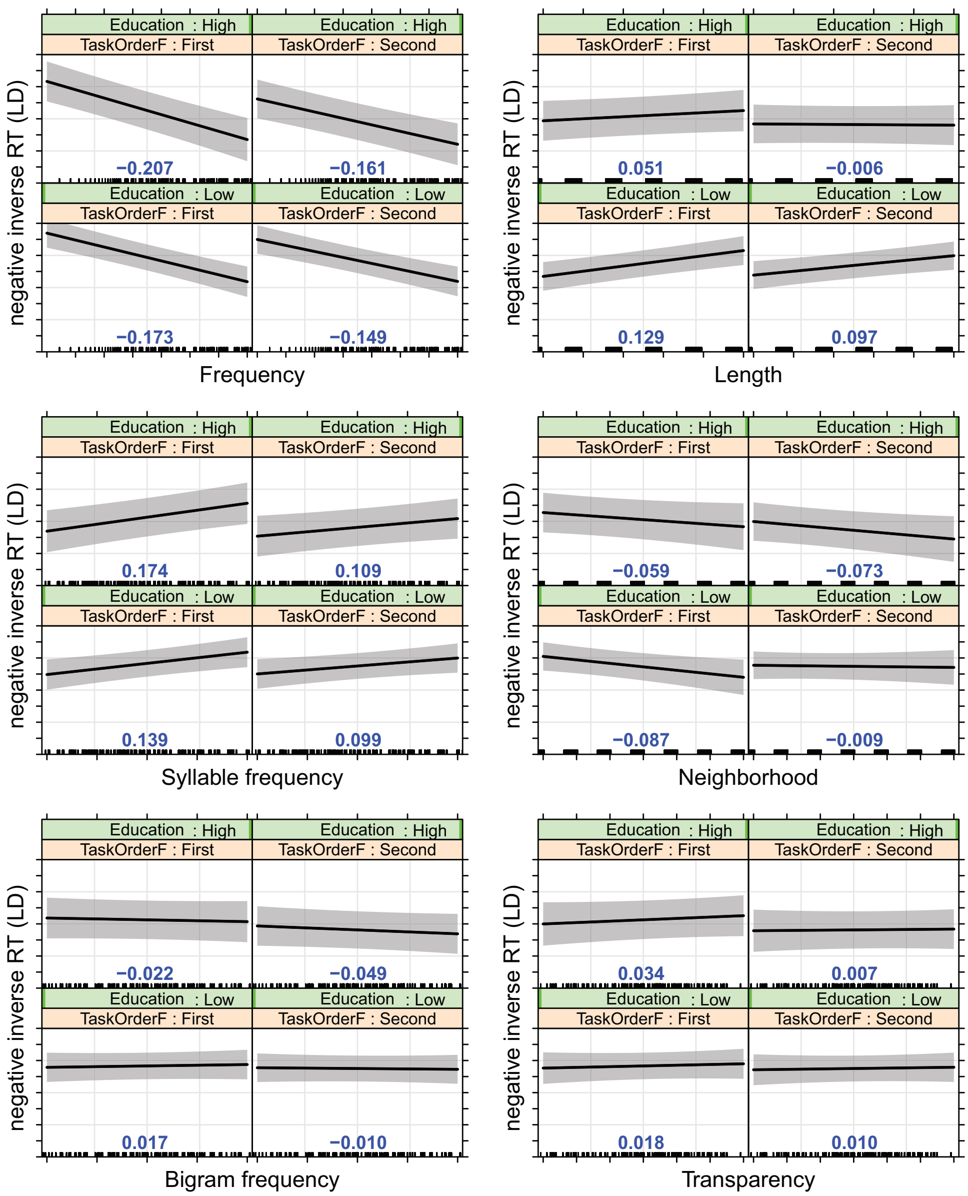

Figure 1. Interaction plots for the six experimental variables and the two modulating factors (education and task order) in lexical decision. All panels extend to the same vertical range (–2.00 to −1.20) and to the full range of the corresponding predictor variable (see variable ranges in Table 1). The vertical range of the predicted variable (thick black line) in each panel is indicated in bold blue (black) numbers. See the online article for the color version of this figure.

Frequency

Significant negative short-term effects on frequency were observed in both tasks, consistent with previous studies of repeated naming and lexical decision using isolated words (Blais et al., 2011; Colombo, Pasini, & Balota, 2006; Visser & Besner, 2001) and with priming effects observed on neologisms (de Vaan, Schreuder, & Baayen, 2007). Visser and Besner (2001) simulated the repetition effect on frequency capitalizing on the decay parameter of the dual-route cascaded model (Coltheart et al., 2001). However, their simulation presented no intervening items between repetitions, rendering this kind of explanation implausible for the delayed-priming type of effect observed in our study. Blais et al. (2011) noted that this effect cannot be attributed to residual activation because lexical activation would lead to a transient effect, soon decaying to baseline. Instead, they suggested that repetition affects the strength of connections between lexical and semantic representations.

An alternative view, in the context of optimal decision models of word recognition (e.g., Norris, 2006; Norris & Kinoshita, 2008, 2012), is that encountering a word in the first task alters the prior

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

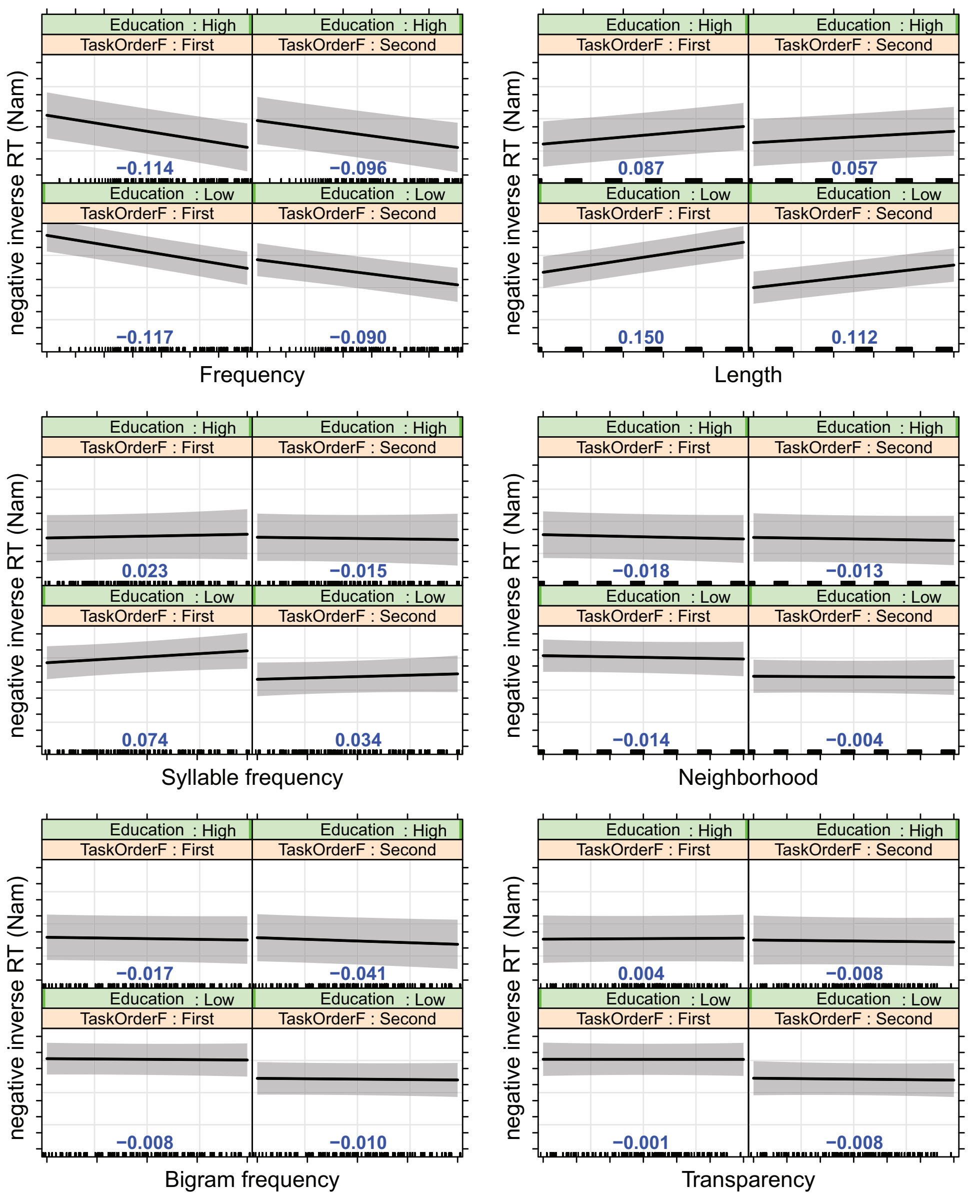

Figure 2. Interaction plots for the six experimental variables and the two modulating factors (education and task order) in naming. All panels extend to the same vertical range (–2.35 to −1.55) and to the full range of the corresponding predictor variable (see variable ranges in Table 1). The vertical range of the predicted variable (thick black line) in each panel is indicated in bold blue (black) numbers. See the online article for the color version of this figure.

probability for this word so that it becomes more likely through the second task. Because low-frequency words have lower priors than high-frequency words (indeed, this is what frequency effects amount to in such models), an update of the prior might affect low-frequency words more than high-frequency words, resulting in reduced frequency effects in the second task. Thus, in these models, short-term frequency effects, such as those observed in our study, can be thought of as part of the normal operation of the model, assuming there is a mechanism for long-term updating of the prior, because the increased expectation for a low-frequency word can only be short-lived. In either a connectionist or optimal-perceiver formulation, a long-term effect of repetition (i.e., cumulative reading experience) can be expected to arise in such situations.

In our data, there was no significant interaction of frequency with education in either task, in contrast to the findings of Tainturier et al. (1992), but in agreement with other reports of null long-term modulation effects with comparable pseudoword sets (i.e., closely matched to the words and including no pseudohomophones; cf. Lewellen et al., 1993; Sears et al., 2008). The null effect cannot be ascribed to inadequate frequency metrics because our experimental variable was derived from a sufficiently large,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

34-million-word printed word corpus (cf. Kuperman & Van Dyke, 2013), likely representative of typical reading experience. However, the correlation between individual frequency effects (random slopes of frequency) and task performance (participant random intercepts) was significant, complicating the interpretation of long-term effects. It appears, then, that education and task performance assess distinct aspects of individual differences that are both relevant for visual word recognition but have differential effects.

Given that frequency effects are overall negative (i.e., longer RTs for lower frequency words), our finding of a significant positive random effects correlation indicates larger (i.e., more negative) frequency effects for faster (i.e., below-average RTs) participants. This is inconsistent with previous studies examining individual differences in verbal skills (Ashby et al., 2005; Chateau & Jared, 2000; Diependaele et al., 2013; Kuperman & Van Dyke, 2011; Schilling et al., 1998; Yap et al., 2012), which have generally reported reduced frequency effects for higher skill participants. As noted in the introduction, some of those findings were qualified by metric or stimulus properties, far from supporting a general conclusion. Moreover, some of the previously reported effects were arguably relative to stable individual differences (traits), such as verbal ability, rather than to cumulative effects of reading experience.

On the other hand, our finding is consistent with reports of larger frequency effects in older than in younger participants that have been ascribed to vocabulary differences (Ramscar et al., 2013, 2014). Far from being specific indices of reading experience, vocabulary skills, as measured by vocabulary tests, are strongly associated with reading comprehension, more so than with word-level reading skills (e.g., in young adults, see Braze, Tabor, Shankweiler, & Mencl, 2007; and in elementary schoolchildren, see Protopapas, Mouzaki, Sideridis, Kotsolakou, & Simos, 2013). Moreover, even though vocabulary correlates with print exposure indices used to assess reading experience (e.g., .73, age-adjusted, between the vocabulary composite and the print experience composite in Braze et al., 2007), it also correlates so highly with IQ (e.g., .85, age-adjusted, in Braze et al., 2007) that it is often used as a proxy for (verbal) cognitive ability. However, the vocabulary size modeled by Ramscar et al. (2013, 2014) is not accurately estimated by standard vocabulary tests, for psychometric reasons: Because of the very small probability of any single very rare word being in one's vocabulary, test items are inadequate probes for the very low frequency range. Thus, "vocabulary" can be an ambiguous and misleading term in this context, and more work will be necessary to disentangle the relationship between vocabulary size, word recognition efficiency, and frequency effects.

Notably, the significant random effects correlation in our data did not arise when conducting the analysis with logarithmically transformed RTs, raising issues of scaling (discussed further below). Assuming, for a moment, that the finding proves replicable, it remains to be determined what aspects of individual differences are and are not related to frequency effects, indexed by processing speed and education, respectively. One might reasonably expect lexical decision and naming speed to correlate with reading experience, but the causal directionality and unique variance associated with the relationship are unknown. If these measures can be considered valid indices of cumulative reading experience, then our data show that frequency effects increase with increasing experience.

On the other hand, years of education seems to be a more closely related (if imperfect) proxy of adult reading experience, to the extent that tertiary education, at least in Greece, seems to consist largely of reading assignments and studying for exams. This suggests that the significant random effects correlation may turn out to be an artifact of inverse RT scaling. In that case, we could conclude that when frequency and RTs are both logarithmically scaled, and when frequency counts are based on a large and representative corpus, there is no difference in the frequency effect among participants who are expert readers, differing in age and education within the range examined in our study, that is, young adults who have at least finished high school and have been admitted to a university program. This conclusion may not be extended to children, less experienced readers, or older adults, who are outside the range of our study. Therefore, from this point of view, it seems that cumulative reading experience does not further alter the frequency effect once reading expertise has been attained.

How can a delayed effect of repetition have no permanent consequences, leaving no trace in an already well-established lexical system? We suggest that short-term effects of repetition do leave permanent traces in the lexical and lexico-semantic connections, but because continued cumulative exposure necessarily reflects the frequencies of lexical items in a sufficiently large lexicon, the *relative* strength structure of connections remains unaffected. In other words, the long-term stability of frequency effects is dynamic, rather than fixed, reflecting the continuous update of connection strengths (or priors, in optimal decision models) at rates consistent with stable item frequencies, once those frequencies have been established in the expert reader through cumulative experience. This proposal, to be verified in computational modeling, highlights the need for representative frequency metrics in word recognition research (cf. Kuperman & Van Dyke, 2013), and it is consistent with a constant updating of expert reading systems, contrasting with existing models that remain fixed once trained (or otherwise initially adjusted). This proposal is also consistent with the findings of Ramscar et al. (2013, 2014)—even though it might appear contradictory—because it concerns the frequency range for which stable proportions are established. Very rare items would not exhibit stable proportions, individually being of very low probability in any given vocabulary. Because the difference in the frequency effect reported by Ramscar et al. concerns the very low frequencies, it is consistent with a prediction of no difference in frequency effects for the remaining spectrum, which is typically sampled in vocabulary tests and in stimulus sets like ours.

Overall, the long-term situation for frequency effects is far from resolved, in our study as well as in the previous literature. Further work should disentangle and clarify the role of factors including aspects of cognitive and verbal ability, overall exposure, and experience with specific items. To this end, principled approaches to the origin of frequency effects (e.g., Baayen, Milin, Đurđević, Hendrix, & Marelli, 2011; Norris, 2006) may prove more helpful than hard-coded parameters (e.g., Coltheart et al., 2001) in permitting computational models to reveal the role of experience-related manipulations.

Length

In contrast to frequency, length effects were consistently modulated by both short-term and long-term variables, and, in addition, there remained a significant correlation with individual task performance beyond the effects of education. These findings are consistent with previous reports (of long-term effects; see Kuperman & Van Dyke, 2011, and Yap et al., 2012) insofar as they may be understood as a general susceptibility of length effects to aspects of verbal and reading efficiency. However, it is not clear how to interpret length effects in word processing from the point of view of either dual-route or optimal decision theories.

On the one hand, the dual-route cascaded model specifically predicts no length effects for words (Coltheart et al., 2001) because length effects arise entirely from the involvement of the nonlexical route, and in the case of real words, the output of the lexical route is available earlier. Maloney, Risko, O'Malley, and Besner (2009) have successfully applied this rationale to investigate the formation of novel lexical items, as a result of repeatedly reading pseudowords, by measuring the progressive reduction of length effects (but cf. Martens & De Jong, 2008). The lack of length effects for words was in agreement with human data available at the time but may be an artifact of restricting the data to monosyllables, as studies including multisyllabic words have reported significant effects of length on naming and lexical decision in English (Yap & Balota, 2009) and other languages (Barca, Burani, & Arduino, 2002; Ferrand et al., 2010, 2011; Keuleers et al., 2010).

Recourse to a nonlexical route might seem like a reasonable candidate explanation for length effects because both overall higher expertise and short-term residual lexical activation may be associated with diminished involvement of a nonlexical procedure, as suggested by Maloney et al. (2009). That is, if the lexical and nonlexical procedures interact such that the nonlexical route can contribute more rapidly to the output for shorter words, consistent with the lexical route, then length effects can emerge in dual architectures. This has apparently been achieved in the connectionist dual process (CDP++) model by greatly increasing the grapheme parsing rate, leading to successful modeling of length effects (Perry et al., 2010). Presumably, a countereffect could be obtained by speeding up lexical activation, either in general or for specific items. Thus, short-term and long-term effects on length may arise from transient and permanent, respectively, speedup of the lexical route, causing the delayed output of the nonlexical route to have less opportunity to affect the outcome.

This approach would predict an interaction of length with frequency because the speed of the sublexical route would be the same for items of the same length, but only the less frequent items would lexically rise slowly enough to be affected. This prediction has been supported in studies reporting a significant interaction between frequency and length (Weekes, 1997; Yap & Balota, 2009); however, it is at odds with Greek data (Protopapas & Kapnoula, 2013). The source of this discrepancy warrants further investigation, as it may be related to the natural correlation between frequency and length in unselected stimuli, in contrast to the stimulus set used here, in which independent variables were decorrelated.

However, length effects need not be ascribed to a nonlexical route. In single-route connectionist models, most existing implementations have only addressed monosyllabic words. Nevertheless, weak length effects emerged in the model of Plaut, McClelland, Seidenberg, and Patterson (1996), the implications of which need to be worked out in more inclusive simulations with multisyllabic items. More recently, Chang, Furber, and Welbourne (2012) ascribed length effects to visual processing and implemented a multilayer connectionist network, in which orthographic representations were learned on the basis of visual input rather than being predefined. The pressure to learn orthographic representations in the additional layer gave rise to effects often attributed to serial processing in dual route frameworks, including word length effects as well as position of irregularity and whammy effects. However, it is not clear whether this formulation could allow emergence of either short-term or long-term effects once the orthographic representation has been learned.

In the revised Bayesian reader (Norris & Kinoshita, 2012), an optimal decision model, length effects were successfully simulated without recourse to serial operations. The origin of length effects in this model has not been scrutinized. Perhaps longer intervals are required for sufficient information to accumulate when the number of letters increases because the identity and position of each individual letter is initially uncertain. Short-term effects on length might be accounted for in this model in the same way as frequency effects: If a word's prior is updated upon initial presentation, then its uncertainty in the second encounter will be reduced. Because longer words are more uncertain, they will be affected more strongly by the update of their prior; hence, length effects will be reduced. Whether this mechanism can also account for long-term effects, given a form of long-term tracking of priors, remains to be determined.

Finally, eye movements may offer a more trivial account for the length effects. For example, in the naïve discriminative reader (Baayen et al., 2011), a fixation penalty was added to simulated latencies for words with more than five letters, accounting for the increased probability of multiple fixations on longer words (Rayner, 1998, 2009). Modeling refixations has also succeeded in producing length effects in a recurrent connectionist network (Plaut, 1999; Plaut, McClelland, & Seidenberg, 1995). Because less skilled readers make more refixations on the same word than more skilled readers (Kuperman & Van Dyke, 2011), this could explain, at least in part, the interaction between length and education in our data, assuming that the increased probability of refixation for less skilled readers would apply more strongly for longer words. In this context, the short-term effect could still be accounted for by the updating of a word's prior probability in the preceding task, which might reduce the need for refixations.

The number of fixations does not seem sufficient to capture the observed effects, however, because our stimuli did not exceed 10 letters in length, arguably being within the identification span of our participants. Moreover, less skilled readers exhibit not only increased probability of refixation but also increased gaze duration (Kuperman & Van Dyke, 2011), consistent with a need for longer processing time beyond moving the eyes more often. At present, the available data cannot logically preclude the possibility that part of our reported length effects may be trivially accounted for by multiple fixations. To test this possibility, future research must (a) track eye movements during these tasks, to document the number of fixations per item and its relationship to length; and (b) repeat the tasks with duration of stimulus presentation not exceeding the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

duration of a fixation, to preclude any benefit of refixation, and examine whether length effects on RTs are still present beyond any effects on accuracy.

Syllable Frequency

Syllable frequency has emerged as an important variable in accounting for item RT variance, consistent with findings in French, German, and Spanish (Conrad, Carreiras, & Jacobs, 2008; Conrad, Carreiras, Tamm, & Jacobs, 2009; Conrad, Grainger, & Jacobs, 2007; Conrad & Jacobs, 2004; Conrad, Tamm, Carreiras, & Jacobs, 2010). In the present study, a long-term modulation of the inhibitory syllable frequency effect approached significance in the naming task, with more years of education associated with a smaller effect. In addition, the syllabic effect was reduced by repetition (a short-term effect) in both tasks.

Similar to the short-term modulation of length effects, this may be attributed to transient residual lexical activation (or perhaps updated priors), resulting in faster lexical processing, causing sublexical influences—including indirect phonological activation—to diminish (cf. Conrad et al., 2007). Similarly, the long-term decrease of the inhibitory syllabic effect (associated with higher education) may reflect a cumulative influence of item activation. In lexical decision, semantic involvement may lead to resonant activation feeding back through the phonological lexicon and negatively impacting recognition via lexical competition. In contrast, in naming, any negative effects of competition might be counteracted by direct *positive* effects in articulatory planning, resulting in a nonsignificant net effect as a combination of two opposing trends, especially for the more experienced readers. Although these are highly speculative accounts, there are currently no alternative proposals for how syllable frequency effects might be accounted for in the context of other types of word recognition models.

Neighborhood

Despite their significance and ubiquity in the literature, neighborhood effects were not a major factor in our data. A marginally significant facilitatory effect of neighborhood size on lexical decision was observed, which, for participants on the low end of our education range, was knocked out by repetition (i.e., a preceding naming task), as indicated by a marginally significant triple interaction. There was no effect on naming. Rather than attempting to interpret a potentially unreliable effect, we might instead point out that the notion of neighborhood is problematic for a database such as ours, in part due to properties of the language such as inflectional morphology and word length distribution.

Specifically, content words in Greek (adjectives, nouns, and verbs) are always inflected by suffixation, for a range of grammatical properties, leading to sets of individual word forms differing minimally at their right end. It is not clear whether members of an inflectional family should count as distinct lexical items and, therefore, as neighbors (i.e., potential competitors) of each other, or whether lexical entries are structured around base forms that may activate each other via their common root. In addition, the preponderance of long and multisyllabic words (median token length in the corpus is five letters, two syllables; Protopapas et al., 2012) results in overall sparse neighborhoods, as there are many more ways for longer strings to be different than shorter strings. The combined effect of these two factors suggests that what is counted as neighbors in our corpus may be largely confined to inflectional variants, especially for the longer words, so that global sparseness is distorted by local denseness. This situation is not unique to Greek, of course, but seems to be quite different from English, which is the most studied language and the one in which the effects of neighbors and their theoretical significance have been established. Further research will be needed to elucidate the notion of neighborhood in languages with longer words and rich inflectional morphology, as far as the crucial properties of lexical activation and competition are concerned.

An interesting possibility was raised by Yarkoni, Balota, and Yap (2008), who suggested that length effects may be neighborhood effects in disguise. Instead of Coltheart's N, they used a neighborhood density metric taking into account the 20 nearest neighbors of each word in terms of edit distance (termed "orthographic Levenshtein distance 20," or "OLD20"). Because they found that the word length effect was greatly diminished by controlling for the OLD20 effect, they proposed that length effects may arise from orthographic similarity (more support due to more neighbors for shorter words) rather than from contributions of a nonlexical route (faster for shorter words). This idea seems compatible with the finding of Chang et al. (2012) that length effects can arise in a feedforward parallel model of word reading at the interface between visual and orthographic representations, insofar as neighborhoods are conceived of as statistical regularities in the orthographic lexicon.

We could not have used OLD20 in our main analysis because this variable has not been decorrelated with the other effects under investigation during item selection. However, to examine this possibility, we added OLD20 to our length analysis as an additive fixed effect, plus a random slope allowed to correlate with the participant intercept and with the length random slope. OLD20 was significantly correlated with number of letters in our word set (Spearman's $\rho = .647$, $p < .001$). Its main effect on RTs was significant in lexical decision ($\beta = 1.15 \times 10^{-1}$, $t = 2.59$, $p = .010$) and marginally in naming ($\beta = 5.73 \times 10^{-2}$, $t = 1.94$, $p = .054$). Addition of OLD20 to the length model reduced the main effect of length from 19.6 ($p = .002$) to a nonsignificant 6.3 ($p = .423$) in lexical decision, and from 27.1 ($p < .001$) to 20.5 ($p < .001$) in naming (these numbers are estimated β multiplied by 1,000, to be comparable with those in Table 3).

Although the complete picture is complicated by significant interactions,[2] the preliminary finding from this analysis seems consistent with the proposal of Yarkoni et al. (2008; and with Ferrand et al., 2010, for French) insofar as naming length effects may arise from articulatory planning demands, whereas lexical decision effects may be a relatively purer indicator of lexical

[2] Task order did not interact with OLD20 in either lexical decision ($\beta = 2.84 \times 10^{-3}$, $t = 0.36$, $p = .719$) or naming ($\beta = 8.03 \times 10^{-3}$, $t = 1.47$, $p = .141$), but persisted in modulating length (lexical decision: $\beta = 4.87 \times 10^{-3}$, $t = 2.70$, $p = .008$; naming: $\beta = 3.42 \times 10^{-3}$, $t = 2.17$, $p = .031$). Education interacted with OLD20 in naming ($\beta = -6.72 \times 10^{-3}$, $t = -2.54$, $p = .011$), but not in lexical decision ($\beta = 1.48 \times 10^{-3}$, $t = 0.39$, $p = .699$), whereas the converse was the case for length (lexical decision: $\beta = -2.68 \times 10^{-3}$, $t = -3.06$, $p = .003$; naming: $\beta = -8.77 \times 10^{-4}$, $t = -1.15$, $p = .252$). Triple interactions were not significant.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

activation. Thus, length effects need not necessarily arise from the involvement of nonlexical processing, as assumed in dual route models, but may turn out to index the orthographic similarity structure of the lexicon, in which longer words typically have fewer similar words to interact with. This may allow future connectionist models to address both neighborhood effects and, thereby, apparent length effects, by using much larger, more inclusive word sets.

Orthographic Variables

Our last two variables of interest, namely, mean bigram frequency and graphophonemic consistency, failed to show consistent effects or interactions. The barely significant interaction of bigram frequency with task order in lexical decision, in the absence of main effects and with estimates of different sign for the two task orders, defies coherent interpretation. This may be due to inappropriate metric selection for these variables. However, it may be noted that bigram frequency effects have not been found to be significant in many previous studies. Balota, Cortese, Sergent-Marshall, Spieler, and Yap (2004) excluded bigram frequency from their reported analyses of English monosyllables because they found them to be unrelated to their dependent measures. In their Footnote 1, they noted "repeated failures to demonstrate an influence of this variable" (p. 296). Similarly, Keuleers et al. (2012) found no effect of this variable in the British Lexicon Project and reported discovering only one previous study reporting an effect. Thus, our study joins previous reports in suggesting that bigram frequency may have been overrated as a variable to control for when matching sets of stimuli (cf. Adelman, 2012).

As for graphophonemic consistency, it seems plausible that the overall high transparency of the Greek orthography may result in consistency ranges that are too small to produce any observable processing effects, especially taking into account the relative influences of much more important variables such as frequency and length. The lack of a long-term modulation of consistency also suggests that sufficient reading expertise is generally attained in the Greek orthography by the end of secondary education to diminish any potential effects arising from graphophonemic complexity. Perhaps more targeted studies using appropriately constructed pseudowords might have more power to reveal effects of orthographic consistency in languages with relatively transparent orthographies.

Implications and Limitations

Coming back to our central question as to the nature of short-term and long-term effects, it seems that the most striking finding lies in the difference between the patterns of frequency and length effects: Both were reduced by a preceding task, but only the latter was modulated by years of education. This may reflect a crucial difference in the nature of these effects and in the way in which exposure to lexical items shapes the underlying processing system. We have already discussed possible mechanisms that may subserve the modulation of frequency and length effects by experience. In the case of frequency, RTs may reflect the relative strength of connections within a stable, dynamically updated word recognition system. In contrast, a plausible explanation for the modulation of length effects by education is that highly expert readers may reach a plateau in terms of reading speed for both short and long words, whereas less experienced readers may rely on less efficient (serial or parallel) processes that are more susceptible to item length. Overall, the critical differences may concern relative versus absolute system dynamics and limiting factors. Computational models of visual word recognition should be able to naturally represent these distinctions to capture the different ways in which gradually accumulating short-term effects transform into persistent individual differences.

As noted in the introduction, most work on word recognition has concentrated on very few languages. In particular, research on short-term and long-term effects is largely dominated by studies in the English language. As it is unclear whether work in an orthographic outlier, such as English, can be fruitfully generalized to other languages and orthographic systems, conversely, it is possible that findings in Greek may not be directly comparable with those of previous studies in English. Although Greek is more similar, in terms of orthographic transparency, to most other European languages than English is, the generalizability of our findings must be empirically established before general conclusions can be drawn.

Having said that, our study has revealed several interesting effects that, if confirmed cross-linguistically, may be used to constrain modeling efforts. For example, approaches that eschew learning, claiming to model the mature (adult) reading system only, rather than its development, might appear less justified in light of the transient and potentially permanent effects of the operation of the mature system on its own properties. In particular, dual-route models would have to allow both transient and lasting modification of their lexical connections on the basis of ongoing activation, whereas connectionist models might need to eschew the absolute distinction between a "training" and a "testing" phase, also allowing ongoing modification of connection strengths beyond the attainment of expert performance. It remains to be seen whether such modifications could produce the observed short-term and long-term modulation of frequency and length effects. Moreover, in either type of model, the relative magnitude of short-term repetition and long-term experience must conform to the respective observed rates, which are consistent with highly unequal power law slopes (Kirsner & Speelman, 1996), perhaps indicating that only a minor proportion of the transient enhancement achieves long-term permanence.

Like connectionist networks, optimal decision models such as the Bayesian reader (Norris, 2006) and the naïve discriminative reader (Baayen et al., 2011) seem directly amenable to the learning manipulations needed to simulate short-term and long-term effects, insofar as these models also learn their cue weights and priors on the basis of input data. Therefore, the learning ability can be extended throughout model operation to permit parameter updating on the basis of experience. It remains to be determined how to balance short-term and long-term effects in order to account for differential effects of current (or recent) context and the cumulative processing history of the model. Related approaches can be found in "rational analysis" formulations, in which current cue weights contribute to activation together with base-level activation, which is determined by a decaying function of previous occurrences, thus reflecting cumulative experience in the form of the log odds that an item will reoccur (Anderson et al., 2004).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

However, there are certain concerns to be taken into account when interpreting our results. First, our methodological approach based on a selected stimulus set with decorrelated properties may have allowed us to disentangle individual variable effects, but it remains to be established that these findings are indeed representative and generalizable. Second, in this study, we have chosen to transform raw RTs via an inverse function, based on examination of the distributions of the RTs and the model residuals. In addition to bringing RT data distributions into reasonably close agreement with the normal distribution, inverse scaling affords a rate interpretation as items per second. On the other hand, logarithmic scaling, which is the option more frequently taken, has the property of naturally accounting for proportional effects, because ratios are transformed to differences, which are then linearly modeled or properly accounted for in random intercepts. However, it is not clear that proportionality is empirically satisfied, or that it is consistent with predictions of information processing models (cf. Faust, Balota, Spieler, & Ferraro, 1999).

Evidently, these choices are not without consequences: Although almost all findings reported above are robust to scaling choice, the correlation between random participant intercepts and random slopes for frequency was significant only with the inverse transformation. It is unclear whether this is a statistical artifact, due to poorly modeled nonlinearities, or a case of enhanced modeling of variance components. Thus, the issue of scaling is complicated by statistical considerations that are contrasted with empirical and modeling considerations. It may take several iterations to clarify both the appropriate scale of analysis as well as the best-fitting curves accounting for the effects of particular variables.

An imperfect aspect of our experimental design was that all items were repeated, rather than having some repeated and some unrepeated items in the second session. In this way, repetition is confounded with session, potentially permitting interpretations of our findings related to generic factors such as fatigue or practice effects. Fatigue seems particularly unlikely as an explanation, because the sizable main effect of task order was negative, that is, lower RTs overall when second than when first. Moreover, it is not clear how generic effects might explain the specific differential effects revealed in our analysis. For example, how could generic practice or fatigue affect longer or low-frequency words more than shorter or high-frequency ones?

If reduced frequency effects were due to practice or fatigue, they should appear within tasks as well; however, the interaction of frequency with trial order was not significant in either task (lexical decision, $\beta = -5.08 \times 10^{-4}$, $t = -0.41$, $p = .683$; naming, $\beta = 1.93 \times 10^{-3}$, $t = 1.09$, $p = .276$) and did not interact with task order (lexical decision, $\beta = 3.07 \times 10^{-4}$, $t = 0.12$, $p = .902$; naming, $\beta = -1.73 \times 10^{-4}$, $t = -0.05$, $p = .961$). Instead, it seems that a more specific effect of practice must be invoked to account for the findings, related to the structure of the orthographic lexicon and to specific properties of the processing system, which is precisely the factor we claim is missing from current word recognition models. Even though the present study presents robust findings based on a large sample and a well-controlled and representative stimulus set, future studies should verify this assertion with an unconfounded repetition manipulation. However, it may be noted that our use of two distinct tasks for the first and second presentation, in addition to lending robustness to the consistent findings, also effectively eliminates a potentially devastating criticism, namely, that short-term effects might be due to experience with the task rather than with the words (cf. Kirsner & Speelman, 1996).

A further limitation concerns our use of education as a proxy for word recognition experience and skill. Additional variables, such as a vocabulary test or an author recognition test, might have been desirable, to quantify more precisely the individual amount of reading experience our participants brought to the task, and hence their likely exposure to the experimental stimuli, as distinct from stable verbal ability traits. However, taking into account that we were interested in effects on the mature reading system, and not arguably developmental differences that may be considered outside the scope of certain word reading models, postsecondary years of education seem like a reasonable candidate, as they must correlate with adult reading experience. In addition, the analysis of random participant variance associated with education and age revealed a substantial effect of education, consistent with our hypothesis. At any rate, the results must be interpreted as they stand on their own: The fact that education significantly modulated the effects of certain variables in a consistent and theoretically interpretable way lends a certain post hoc support to the use of this particular proxy.

Finally, it should be taken into account that, in order to isolate and discuss the effects of individual variables, we have ignored interactions among variables, some of which are significant in this data set (see Protopapas & Kapnoula, 2013), and that, in order to examine the correlations between overall performance and individual-level effects, we have greatly simplified the random structure of our models (with the additional nonnegligible benefit of facilitating convergence). Although examination of the full model gives no cause for concern over the implications of these limitations, it remains the case that future studies with larger data sets may be able to take more inclusive approaches to statistical modeling.

Conclusion

In conclusion, we have examined a set of lexical and sublexical variables in accounting for lexical decision and naming latency variance. In particular, we investigated how the effects of these variables are modulated by short-term and long-term individual differences, namely, the experimental manipulation of repetition and the educational level of the participants, respectively. We have revealed differential modulation of the effects of the experimental variables, with word length effects being influenced both by short-term and long-term factors; in contrast, frequency was clearly affected by the short-term factor, but the long-term findings were inconsistent. Syllabic frequency was an important predictor, subject to differential short-term and long-term effects in the two tasks. Potential explanations were offered for these effects, in the context of current modeling frameworks, all of which are, in their current implementations, too static to accommodate such effects. Thus, these findings may be useful to further our understanding of the visual word processing architecture and to constrain computational modeling efforts toward continuous dynamic self-adjustment.

References

Adelman, J. S. (2012). Methodological issues with words. In J. S. Adelman (Ed.), *Visual word recognition* (Vol. 1, pp. 116–138). Hove, UK: Psychology Press.

Adelman, J. S., & Brown, G. D. (2008a). Methods of testing and diagnosing model error: Dual and single route cascaded models of reading aloud. *Journal of Memory and Language, 59,* 524–544. http://dx.doi.org/10.1016/j.jml.2007.11.008

Adelman, J. S., & Brown, G. D. (2008b). Modeling lexical decision: The form of frequency and diversity effects. *Psychological Review, 115,* 214–229. http://dx.doi.org/10.1037/0033-295X.115.1.214

Adelman, J. S., Marquis, S. J., Sabatos-De Vito, M. G., & Estes, Z. (2013). The unexplained nature of reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1037–1053. http://dx.doi.org/10.1037/a0031829

Adelman, J. S., Sabatos-DeVito, M. G., Marquis, S. J., & Estes, Z. (2014). Individual differences in reading aloud: A mega-study, item effects, and some models. *Cognitive Psychology, 68,* 113–160. http://dx.doi.org/10.1016/j.cogpsych.2013.11.001

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111,* 1036–1060. http://dx.doi.org/10.1037/0033-295X.111.4.1036

Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General, 139,* 299–318. http://dx.doi.org/10.1037/a0018366

Andrews, S., & Lo, S. (2012). Not all skilled readers have cracked the code: Individual differences in masked form priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 152–163. http://dx.doi.org/10.1037/a0024953

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 58,* 1065–1086. http://dx.doi.org/10.1080/02724980443000476

Baayen, R. H. (2008). *Analyzing linguistic data*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511801686

Baayen, R. H. (2013). Multivariate statistics. In R. Podesva & D. Sharma (Eds.), *Research methods in linguistics* (pp. 337–372). New York, NY: 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. http://dx.doi.org/10.1016/j.jml.2007.12.005

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. http://dx.doi.org/10.1016/j.jml.2006.03.008

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3,* 12–28.

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review, 118,* 438–481. http://dx.doi.org/10.1037/a0023851

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133,* 283–316. http://dx.doi.org/10.1037/0096-3445.133.2.283

Balota, D. A., Yap, M. J., & Cortese, M. J. (2006). Visual word recognition: The journey from features to meaning (a travel update). In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 285–375). Amsterdam, Netherlands: Academic Press. http://dx.doi.org/10.1016/B978-012369374-7/50010-9

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39,* 445–459. http://dx.doi.org/10.3758/BF03193014

Balota, D. A., Yap, M. J., Hutchison, K., & Cortese, M. J. (2012). Megastudies: What do millions (or so) of trials tell us about lexical processing? In J. S. Adelman (Ed.), *Visual word recognition* (Vol. 1, pp. 90–115). Hove, UK: Psychology Press.

Barca, L., Burani, C., & Arduino, L. S. (2002). Word naming times and psycholinguistic norms for Italian nouns. *Behavior Research Methods, Instruments & Computers, 34,* 424–434. http://dx.doi.org/10.3758/BF03195471

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1. 0–5. Retrieved from http://CRAN.R-project.org/package=lme4

Blais, C., & Besner, D. (2007). Reading aloud: When the effect of stimulus quality distinguishes between cascaded and thresholded components. *Experimental Psychology, 54,* 215–224. http://dx.doi.org/10.1027/1618-3169.54.3.215

Blais, C., O'Malley, S., & Besner, D. (2011). On the joint effects of repetition and stimulus quality in lexical decision: Looking to the past for a new way forward. *The Quarterly Journal of Experimental Psychology, 64,* 2368–2382. http://dx.doi.org/10.1080/17470218.2011.591535

Braze, D., Tabor, W., Shankweiler, D. P., & Mencl, W. E. (2007). Speaking up for vocabulary: Reading skill differences in young adults. *Journal of Learning Disabilities, 40,* 226–243. http://dx.doi.org/10.1177/00222194070400030401

Chang, Y. N., Furber, S., & Welbourne, S. (2012). "Serial" effects in parallel models of reading. *Cognitive Psychology, 64,* 267–291. http://dx.doi.org/10.1016/j.cogpsych.2012.01.002

Chateau, D., & Jared, D. (2000). Exposure to print and word recognition processes. *Memory & Cognition, 28,* 143–153. http://dx.doi.org/10.3758/BF03211582

Coltheart, M., Davelaar, E., Jonasson, T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention & performance IV* (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. http://dx.doi.org/10.1037/0033-295X.108.1.204

Colombo, L., Pasini, M., & Balota, D. A. (2006). Dissociating the influence of familiarity and meaningfulness from word frequency in naming and lexical decision performance. *Memory & Cognition, 34,* 1312–1324.

Conrad, M., Carreiras, M., & Jacobs, A. M. (2008). Contrasting effects of token and type syllable frequency in lexical decision. *Language and Cognitive Processes, 23,* 296–326. http://dx.doi.org/10.1080/01690960701571570

Conrad, M., Carreiras, M., Tamm, S., & Jacobs, A. M. (2009). Syllables and bigrams: Orthographic redundancy and syllabic units affect visual word recognition at different processing levels. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 461–479. http://dx.doi.org/10.1037/a0013480

Conrad, M., Grainger, J., & Jacobs, A. M. (2007). Phonology as the source of syllable frequency effects in visual word recognition: Evidence from French. *Memory & Cognition, 35,* 974–983. http://dx.doi.org/10.3758/BF03193470

Conrad, M., & Jacobs, A. (2004). Replicating syllable frequency effects in Spanish in German: One more challenge to computational models of visual word recognition. *Language and Cognitive Processes, 19,* 369–390. http://dx.doi.org/10.1080/01690960344000224

Conrad, M., Tamm, S., Carreiras, M., & Jacobs, A. M. (2010). Simulating syllable frequency effects within an interactive activation framework.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

European Journal of Cognitive Psychology, 22, 861–893. http://dx.doi.org/10.1080/09541440903356777

de Vaan, L., Schreuder, R., & Baayen, R. H. (2007). Regular morphologically complex neologisms leave detectable traces in the mental lexicon. *The Mental Lexicon, 2,* 1–23. http://dx.doi.org/10.1075/ml.2.1.02vaa

Diependaele, K., Brysbaert, M., & Neri, P. (2012). How noisy is lexical decision? *Frontiers in Psychology, 3,* 348. http://dx.doi.org/10.3389/fpsyg.2012.00348

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *The Quarterly Journal of Experimental Psychology, 66,* 843–863. http://dx.doi.org/10.1080/17470218.2012.720994

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin, 125,* 777–799. http://dx.doi.org/10.1037/0033-2909.125.6.777

Ferrand, L., Brysbaert, M., Keuleers, E., New, B., Bonin, P., Méot, A., . . . Pallier, C. (2011). Comparing word processing times in naming, lexical decision, and progressive demasking: Evidence from chronolex. *Frontiers in Psychology, 2,* 306. http://dx.doi.org/10.3389/fpsyg.2011.00306

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., . . . Pallier, C. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods, 42,* 488–496. http://dx.doi.org/10.3758/BRM.42.2.488

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

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software, 8,* 1–27.

Harm, M. W., McCandliss, B. D., & Seidenberg, M. S. (2003). Modeling the successes and failures of interventions for disabled readers. *Scientific Studies of Reading, 7,* 155–182. http://dx.doi.org/10.1207/S1532799XSSR0702_3

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106,* 491–528. http://dx.doi.org/10.1037/0033-295X.106.3.491

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111,* 662–720. http://dx.doi.org/10.1037/0033-295X.111.3.662

Katz, L., Lee, C. H., Tabor, W., Frost, S. J., Mencl, W. E., Sandak, R., . . . Pugh, K. R. (2005). Behavioral and neurobiological effects of printed word repetition in lexical decision and naming. *Neuropsychologia, 43,* 2068–2083. http://dx.doi.org/10.1016/j.neuropsychologia.2005.03.022

Keuleers, E. (2013). *vwr: Useful functions for visual word recognition research*. R package version 0. 3.0. Retrieved from http://CRAN.R-project.org/package=vwr

Keuleers, E., & Brysbaert, M. (2011). Detecting inherent bias in lexical decision experiments with the LD1NN algorithm. *The Mental Lexicon, 6,* 34–52. http://dx.doi.org/10.1075/ml.6.1.02keu

Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology, 1,* 174. http://dx.doi.org/10.3389/fpsyg.2010.00174

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods, 44,* 287–304. http://dx.doi.org/10.3758/s13428-011-0118-4

Kirsner, K., & Speelman, C. (1996). Skill acquisition and repetition priming: One principle, many processes? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 563–575. http://dx.doi.org/10.1037/0278-7393.22.3.563

Kuperman, V., Drieghe, D., Keuleers, E., & Brysbaert, M. (2013). How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora and megastudies. *The Quarterly Journal of Experimental Psychology, 66,* 563–580. http://dx.doi.org/10.1080/17470218.2012.658820

Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language, 65,* 42–73. http://dx.doi.org/10.1016/j.jml.2011.03.002

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 802–823. http://dx.doi.org/10.1037/a0030859

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2013). *lmerTest: Tests for random and fixed effects for linear mixed effect models*. R package version 2. 0–3. Retrieved from http://CRAN.R-project.org/package=lmerTest

Lemoine, H. E., Levy, B. A., & Hutchinson, A. (1993). Increasing the naming speed of poor readers: Representations formed across repetitions. *Journal of Experimental Child Psychology, 55,* 297–328. http://dx.doi.org/10.1006/jecp.1993.1018

Levy, B. A., Di Persio, R., & Hollingshead, A. (1992). Fluent rereading: Repetition, automaticity, and discrepancy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 957–971. http://dx.doi.org/10.1037/0278-7393.18.5.957

Levy, B. A., Nicholls, A., & Kohen, D. (1993). Repeated readings: Process benefits for good and poor readers. *Journal of Experimental Child Psychology, 56,* 303–327. http://dx.doi.org/10.1006/jecp.1993.1037

Lewellen, M. J., Goldinger, S. D., Pisoni, D. B., & Greene, B. G. (1993). Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision, and semantic categorization. *Journal of Experimental Psychology: General, 122,* 316–330. http://dx.doi.org/10.1037/0096-3445.122.3.316

Maloney, E., Risko, E. F., O'Malley, S., & Besner, D. (2009). Tracking the transition from sublexical to lexical processing: On the creation of orthographic and phonological lexical representations. *The Quarterly Journal of Experimental Psychology, 62,* 858–867. http://dx.doi.org/10.1080/17470210802578385

Martens, V. E., & De Jong, P. F. (2008). Effects of repeated reading on the length effect in word and pseudoword reading. *Journal of Research in Reading, 31,* 40–54. http://dx.doi.org/10.1111/j.1467-9817.2007.00360.x

Mulatti, C., Peressotti, F., Job, R., Saunders, S., & Coltheart, M. (2012). Reading aloud: The cumulative lexical interference effect. *Psychonomic Bulletin & Review, 19,* 662–667. http://dx.doi.org/10.3758/s13423-012-0269-z

Mulatti, C., Reynolds, M. G., & Besner, D. (2006). Neighborhood effects in reading aloud: New findings and new challenges for computational models. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 799–810. http://dx.doi.org/10.1037/0096-1523.32.4.799

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review, 113,* 327–357. http://dx.doi.org/10.1037/0033-295X.113.2.327

Norris, D., & Kinoshita, S. (2008). Perception as evidence accumulation and Bayesian inference: Insights from masked priming. *Journal of Experimental Psychology: General, 137,* 434–455. http://dx.doi.org/10.1037/a0012799

Norris, D., & Kinoshita, S. (2012). Reading through a noisy channel: Why there's nothing special about the perception of orthography. *Psychological Review, 119,* 517–545. http://dx.doi.org/10.1037/a0028450

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. http://dx.doi.org/10.1037/0033-295X.114.2.273

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology, 61,* 106–151. http://dx.doi.org/10.1016/j.cogpsych.2010.04.001

Plaut, D. C. (1999). A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. *Cognitive Science, 23,* 543–568. http://dx.doi.org/10.1207/s15516709cog2304_7

Plaut, D. C. (2006). Connectionist approaches to reading. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 24–38). Oxford, UK: Blackwell.

Plaut, D. C., McClelland, J. L., & Seidenberg, M. S. (1995). Reading exception words and pseudowords: Are two routes really necessary? In J. P. Levy, D. Bairaktaris, J. A. Bullinaria, & P. Cairns (Eds.), *Connectionist models of memory and language* (pp. 145–160). London, UK: UCL Press.

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,* 56–115. http://dx.doi.org/10.1037/0033-295X.103.1.56

Protopapas, A. (2007). CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods, 39,* 859–862. http://dx.doi.org/10.3758/BF03192979

Protopapas, A., & Kapnoula, E. C. (2013). Exploring word recognition with selected stimuli: The case for decorrelated parameters. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1169–1174). Austin, TX: Cognitive Science Society.

Protopapas, A., Mouzaki, A., Sideridis, G. D., Kotsolakou, A., & Simos, P. G. (2013). The role of vocabulary in the context of the simple view of reading. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 29,* 168–202. http://dx.doi.org/10.1080/10573569.2013.758569

Protopapas, A., Tzakosta, M., Chalamandaris, A., & Tsiakoulis, P. (2012). IPLR: An online resource for Greek word-level and sublexical information. *Language Resources and Evaluation, 46,* 449–459. http://dx.doi.org/10.1007/s10579-010-9130-z

Protopapas, A., & Vlahou, E. L. (2009). A comparative quantitative analysis of Greek orthographic transparency. *Behavior Research Methods, Instruments & Computers, 41,* 991–1008. http://dx.doi.org/10.3758/BRM.41.4.991

Ramscar, M., Hendrix, P., Love, B., & Baayen, R. H. (2013). Learning is not decline: The mental lexicon as a window into cognition across the lifespan. *The Mental Lexicon, 8,* 450–481. http://dx.doi.org/10.1075/ml.8.3.08ram

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science, 6,* 5–42. http://dx.doi.org/10.1111/tops.12078

Raney, G. E., & Rayner, K. (1995). Word frequency effects and eye movements during two readings of a text. *Canadian Journal of Experimental Psychology, 49,* 151–173. http://dx.doi.org/10.1037/1196-1961.49.2.151

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422. http://dx.doi.org/10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62,* 1457–1506. http://dx.doi.org/10.1080/17470210902816461

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Rey, A., Courrieu, P., Schmidt-Weigand, F., & Jacobs, A. M. (2009). Item performance in visual word recognition. *Psychonomic Bulletin & Review, 16,* 600–608. http://dx.doi.org/10.3758/PBR.16.3.600

Reynolds, M., & Besner, D. (2002). Neighbourhood density effects in reading aloud: New insights from simulations with the DRC model. *Canadian Journal of Experimental Psychology, 56,* 310–318. http://dx.doi.org/10.1037/h0087407

Reynolds, M., & Besner, D. (2004). Neighbourhood density, word frequency, and spelling-sound regularity effects in naming: Similarities and differences between skilled readers and the Dual Route Cascaded Computational model. *Canadian Journal of Experimental Psychology, 58,* 13–31. http://dx.doi.org/10.1037/h0087437

Schilling, H. E., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition, 26,* 1270–1281.

Sears, C. R., Siakaluk, P. D., Chow, V. C., & Buchanan, L. (2008). Is there an effect of print exposure on the word frequency effect and the neighborhood size effect? *Journal of Psycholinguistic Research, 37,* 269–291. http://dx.doi.org/10.1007/s10936-008-9071-5

Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94,* 143–174. http://dx.doi.org/10.1348/000712603321661859

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin, 134,* 584–615. http://dx.doi.org/10.1037/0033-2909.134.4.584

Spencer, K. A. (2009). Feedforward, -backward, and neutral transparency measures for British English. *Behavior Research Methods, 41,* 220–227. http://dx.doi.org/10.3758/BRM.41.1.220

Tainturier, M. J., Tremblay, M., & Lecours, A. R. (1992). Educational level and the word frequency effect: A lexical decision investigation. *Brain and Language, 43,* 460–474. http://dx.doi.org/10.1016/0093-934X(92)90112-R

Taylor, T. E., & Lupker, S. J. (2001). Sequential effects in naming: A time-criterion account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 117–138. http://dx.doi.org/10.1037/0278-7393.27.1.117

Visser, T. A., & Besner, D. (2001). On the dominance of whole-word knowledge in reading aloud. *Psychonomic Bulletin & Review, 8,* 560–567. http://dx.doi.org/10.3758/BF03196192

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 50A,* 439–456. http://dx.doi.org/10.1080/713755710

Wurm, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does *not* do). *Journal of Memory and Language, 72,* 37–48. http://dx.doi.org/10.1016/j.jml.2013.12.003

Yap, M., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language, 60,* 502–529. http://dx.doi.org/10.1016/j.jml.2009.02.001

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance, 38,* 53–79. http://dx.doi.org/10.1037/a0024177

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15,* 971–979. http://dx.doi.org/10.3758/PBR.15.5.971

Ziegler, J. C. (2011). Understanding developmental dyslexia through computational modeling: An individual deficit-based simulation approach. In P. McCardle, B. Miller, J. R. Lee, & O. J. L. Tzeng (Eds.), *Dyslexia across languages: Orthography and the brain-gene-behavior link* (pp. 169–183). Baltimore, MD: Paul Brookes.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Ziegler, J. C., Castel, C., Pech-Georgel, C., George, F., Alario, F.-X., & Perry, C. (2008). Developmental dyslexia and the dual route model of reading: Simulating individual differences and subtypes. *Cognition, 107,* 151–178. http://dx.doi.org/10.1016/j.cognition.2007.09.004

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. http://dx.doi.org/10.1037/0033-2909.131.1.3

Ziegler, J. C., & Goswami, U. (2006). Becoming literate in different languages: Similar problems, different solutions. *Developmental Science, 9,* 429–436. http://dx.doi.org/10.1111/j.1467-7687.2006.00509.x

Ziegler, J. C., Perry, C., & Zorzi, M. (2014). Modelling reading development through phonological decoding and self-teaching: Implications for dyslexia. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 369,* 20120397. http://dx.doi.org/10.1098/rstb.2012.0397

Appendix A

Benefits of Decorrelating Variables

In the following simulations, we demonstrate that using samples selected to decorrelate variables that are correlated in the population results in (a) more accurate estimates of the effects of each variable when estimated individually, and (b) reduced susceptibility to interaction artifacts when nonlinear effects are modeled linearly. In both cases, the advantage of decorrelated samples increases with the magnitude of the correlation in the population.

In the simulations we created "populations" of $N = 200{,}000$ value pairs, corresponding to variables x_1 and x_2, sampled randomly from bivariate normal distributions with variance-covariance matrix $\begin{bmatrix} 1 & c \\ c & 1/2 \end{bmatrix}$, with c ranging between 0 and 0.25 in steps of 0.01. Two sets of $n = 200$ (x_1, x_2) pairs were sampled from this population. One set was sampled completely at random (within $-2.5 < z < 2.5$ of the scaled values, to avoid extremes). The second set was sampled repeatedly (within a scaled radius of 2.5 from the mean) until the Spearman correlation between the two variables satisfied $|\rho| < 0.1$. Thus, we obtained one random and

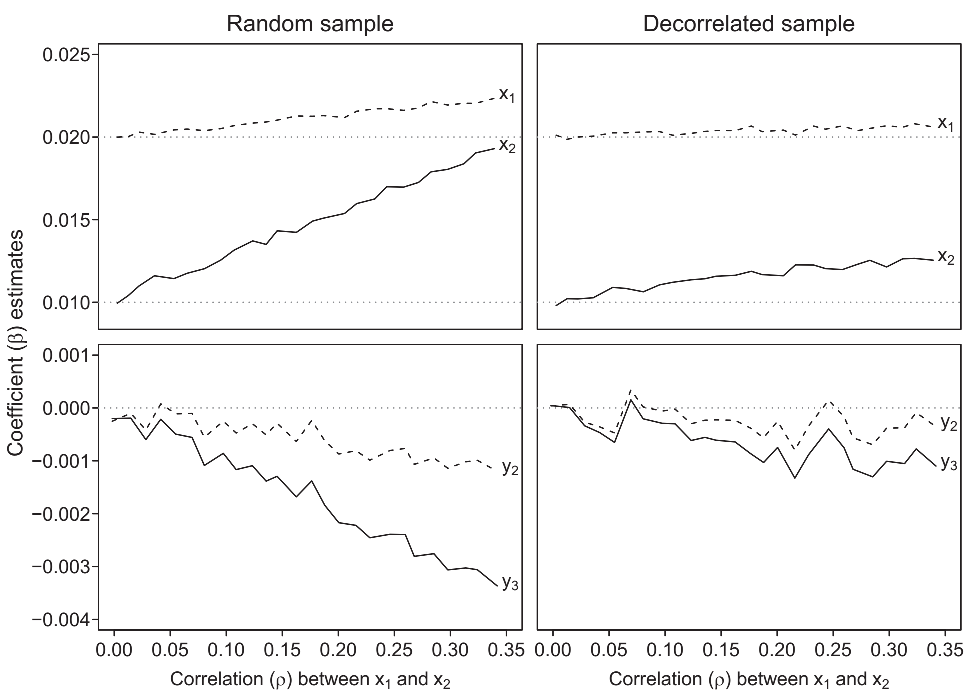

Figure A1. Estimated regression coefficients for the effects of x_1 and x_2 (independently, in simple models; top) and their interaction (in full models of y_2 and y_3; bottom) in the random sample (left) and the decorrelated sample (right) as a function of the correlation (Spearman's ρ) between x_1 and x_2 in the population. Dotted lines mark the correct values.

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

one artificially decorrelated sample of size 1/1,000 of the population. This was repeated 1,000 times for each step of c.

In the first simulation, a "dependent" variable y_1 was defined as $y_1 = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \varepsilon$, with $b_0 = 0.00$, $b_1 = 0.02$, $b_2 = 0.01$, and $\varepsilon = 0.05 \cdot N\ (0,1)$. That is, this dependent variable included additive linear effects of both independent variables (larger for the higher variance variable) plus random noise that was 5 times greater than the weaker of the two effects. The combined effects of x_1 and x_2 accounted for approximately 15% of y_1 variance (multiple R^2) in the population. Subsequently, linear models were fit, separately for the effects of each independent variable, using the random and the decorrelated sample. Figure A1 (top) shows the estimated coefficients for the effects of x_1 and x_2 as a function of the correlation between x_1 and x_2 in the population. The correct values (0.2 and 0.1, respectively) are marked with a dotted line. Clearly, the random sample (left panel) overestimates these effects, particularly of the (lower-variance) x_2, as a result of the correlation, whereas the decorrelated sample (right panel) is much less susceptible to this artifact. This simulation demonstrates that using decorrelated samples is effective in permitting reasonably accurate individual analyses of the decorrelated variables.

In the second simulation, dependent variables y_2 and y_3 were defined similarly to y_1, except that the effect of x_1 was nonlinear. Specifically, $b_1 \cdot x_1$ was replaced by $b_1 \cdot x'$, where x' was nonlinearly derived from x_1 and then scaled to $M = 0$ and $SD = 1$. Two nonlinearities were tested. The first one (for y_2) was a very mild logarithmic, in which $x' = \log_{10}(x_1+6)$. The second one (for y_3) was a moderate quadratic, with $x' = -(x_1 - 2)^2$. Figure A2 displays these nonlinearities for $c = 0$. Therefore, in each version of this simulation, the dependent variable included one linear and one nonlinear effect of the independent variables but no interaction between the two. Linear models were fit, separately for each version of the dependent variable, including linear effects of the two independent variables as well as an interaction term. Figure A1 (bottom) shows the estimated coefficient for the interaction, for each dependent variable, as a function of the correlation between x_1 and x_2 in the population. The correct value (0.0) is marked with a dotted line. An artifactual interaction is evident in the random sample, especially in the case of the quadratic effect (solid line), as a result of the correlation and the misspecified model (which includes only a linear effect), whereas the decorrelated sample is much less susceptible to this artifact. This simulation demonstrates that using decorrelated samples is effective in isolating nonadditivities due to interactions from nonadditivities due to nonlinearities, thereby permitting more accurate estimation of interaction effects in the presence of unmodeled nonlinearities.

The results of the two simulations are consistent with the claim that using samples with decorrelated variables can help estimate more accurately the effects of multiple variables that are naturally correlated in the population, and in this way can produce complementary and corroborating evidence that is less susceptible to statistical artifacts arising from the correlation than are randomly selected samples. The extent to which decorrelated samples can be useful in practice depends on the correlations in the population and on the presence of unmodeled nonlinearities.

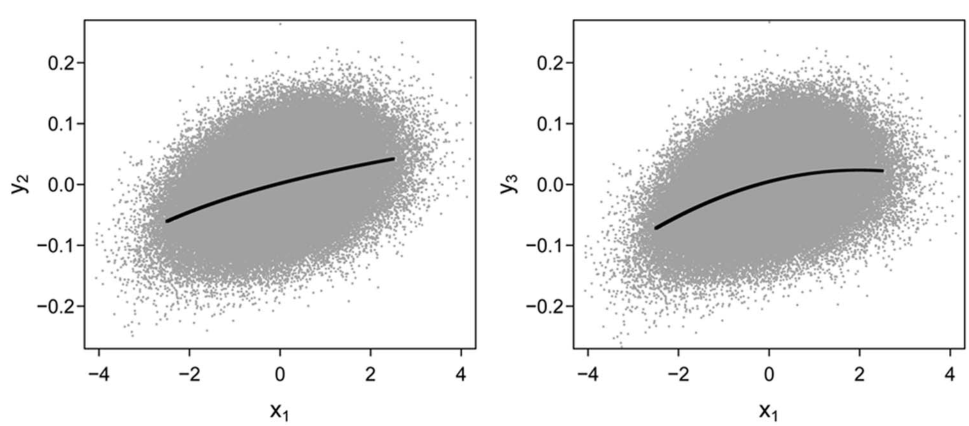

Figure A2. Distribution of y_2 and y_3 as a function of x_1 in a population with zero covariance between x_1 and x_2 (gray points). The black line marks the denoised trend, to illustrate the logarithmic (left) and quadratic (right) relationship.

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix B

Distributions of Experimental Variables

Figure B1 displays the distributions of the six experimental variables in the stimulus sample compared with their distribution in the corpus they were derived from. The top five rows display the low triangle of the bivariate distributions array, in which corpus items (i.e., individual word forms) are marked with light gray dots, and the stimulus set is marked with black bullets on top of the corpus background. The bottom row shows the density plots of the stimulus set (continuous line) against the density plots of the corpus type distribution (dashed line) and token distribution (dotted line). The intercorrelations among all measured variables in the corpus are listed in Table B1.

Although the sampled variables span wide ranges, including the most densely populated ranges in the corpus, the issue of repre-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

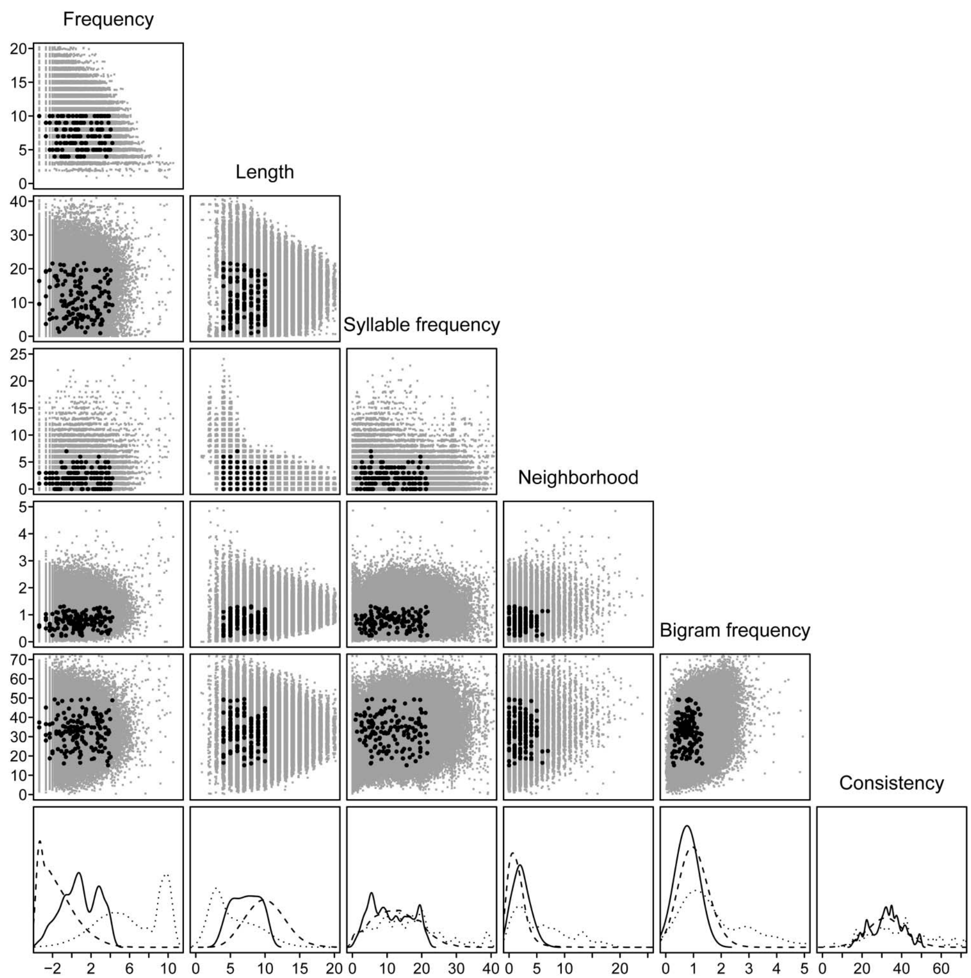

Figure B1. Top five rows: Bivariate distributions of the six experimental variables in the stimulus sample (black dots) overlaid on their distribution in the corpus (jittered; gray dots). Bottom row: Density plots of the same variables in the stimulus set (continuous line) against corpus type (dashed line) and token (dotted line) distribution.

(*Appendices continue*)

Table B1
Nonparametric Correlation Coefficients (Spearman's ρ) Between Variables for Corpus Types and Tokens

	1	2	3	4	5	6	7	8	9	10	11
1. Log frequency[a]		−.220	−.220	−.180	.260	.280	.040	.050	.040	.070	.020
2. Number of letters[a]	−.840		.950	.870	−.320	−.430	.200	.080	.120	.070	.050
3. Number of phonemes	−.860	.970		.890	−.280	−.430	.100	.180	.150	.080	.160
4. Number of syllables	−.840	.930	.940		−.250	−.370	.140	.200	.370	.360	.180
5. Orth. neighbors[a]	.740	−.800	−.800	−.780		.670	.040	.110	.020	.030	.070
6. Phon. neighbors	.700	−.770	−.770	−.740	.800		.060	.100	.000	.050	.030
7. Orth. bigram frequency[a]	.590	−.410	−.460	−.510	.540	.520		.440	.200	.150	.430
8. Phon. bigram frequency	.440	−.320	−.270	−.370	.380	.510	.690		.220	.320	.560
9. Orth. syllable frequency	.270	−.220	−.210	.000	.210	.190	.150	.100		.850	.360
10. Phon. syllable frequency[a]	.180	−.180	−.190	.070	.100	.120	.000	.010	.810		.260
11. G-P consistency[a]	.350	−.410	−.290	−.340	.320	.320	.390	.540	.230	.050	

Note. Coefficients for corpus types are above the diagonal; those for tokens are below the diagonal. Corpus types, $N = 206{,}621$; tokens, $N = 29{,}557{,}090$. Orth. = orthographic; Phon. = phonological; G-P = graphophonemic.
[a] Variable in the "experimental" set.

sentativeness is not easy to resolve. What is representative with respect to the flat lexicon (word type distribution) is not representative with respect to any participant's experience (word token distribution). That is, if we sample every word with equal probability, rare words will be overrepresented, potentially obscuring typical lexical processing (i.e., of more frequent items) as well as decreasing reliability (because RTs to low-frequency words are less reliable; Diependaele, Brysbaert, & Neri, 2012). Conversely, if we sample based on probability of occurrence, rare words will be underrepresented, potentially distorting the general properties of the processing system. It is unclear which kind of sampling will best reveal the properties of lexical processing and their behavioral effects on visual word recognition. Thus, in the present approach, we have taken a middle ground, by ensuring adequate coverage of variable ranges and by seeking to achieve sample distributions that are intermediate between type and token distributions in the corpus.

The stimulus selection process was initiated with 150 words and 150 pseudowords that were handpicked to span wide ranges in the experimental variables. Words were selected from the IPLR C corpus (Protopapas et al., 2012), and pseudowords were selected from a set of approximately 1,000 pseudowords constructed specifically for this study. Nonparametric correlation coefficients (Spearman's ρ) between all variables were calculated, for each of the two sets separately, and items were replaced on the basis of the correlations. For example, if frequency showed a high positive correlation with neighborhood size, then items with high frequency and high neighborhood size, or with low frequency and low neighborhood size, were removed from the set. Items with high frequency and low neighborhood size were then sought in the corresponding corpus, taking into account the other pairwise correlations as much as possible. This process was repeated multiple times until all qualitatively distinct variables were not significantly correlated.

Initial phoneme classes were not among the decorrelated variables and were only expected (and confirmed) to affect naming times (not lexical decision). To examine possible associations of initial phonemes with the experimental variables, one-way analyses of variance were conducted with each experimental variable as the dependent variable, separately for first phoneme class and for second phoneme class as the independent variable. Table B2 lists the results of these analyses. Naturally, phoneme classes were associated with syllable frequency. Associations with other variables would not survive correction for multiple comparisons.

Table B2
ANOVA of Each Experimental Variable by First and Second Phoneme Over the Word Set

	First phoneme class		Second phoneme class	
Dependent variable	$F(4, 145)$	p	$F(5, 144)$	p
Frequency	.21	.932	1.53	.184
Length	.71	.584	.70	.624
Syllable frequency	42.00	<.001	13.05	<.001
Neighborhood	.35	.847	.45	.815
Bigram frequency	2.98	.021	1.41	.224
Consistency	1.50	.205	2.42	.039

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix C

List of Stimuli

Words

αδερφές, άθλησης, άθραυστος, αλφαβήτου, ανδρισμός, ανθρακικής, αντιληπτά, αρχέγονο, άσμα, ασφαλείς, αυθαίρετα, αυξάνουν, άφθαρτος, βαρύτερα, βασάνιζε, βγαίναμε, βόλεμα, βρεγμένο, γλιτώσαμε, γνώριμοι, γόης, γόπα, γύρισμα, δάσκαλος, δεσπότη, δήλωσαν, δίσκο, διχασμένο, δολοφονήσω, δώστε, εβδομάδας, εκδικητής, εκπροσώπου, έλαβε, ελεούσα, έλκη, εμπρός, ενεργή, επέβαλε, ερμηνευθεί, εύρος, εφτά, εχθρούς, εχτές, ζακέτα, ζορίζεται, ηγεμονίας, ημεδαπό, ήπαρ, ηχητικός, θεάνθρωπος, θέμιδα, θερμικές, ινδικό, ισόπαλα, ισοφάρισε, καθαρόαιμη, καθρέφτες, καλπάζει, καρπός, κοστίζει, κόψη, λαβύρινθος, λαούς, λάσπη, μαχόμενη, μεμπτές, μεταβίβασα, μέχρις, μισθωτοί, μνηστή, μονάδα, νάρκη, ναύλος, μέθη, νέφος, νεφρική, νησιά, νυχτερινές, ξεναγός, ξενύχτησα, ξιφία, οθόνη, ολόκληρα, όμματα, ομόκεντρο, οπλές, οσμή, παλλαϊκές, παλμός, πάπιες, παράγραφος, παράδοξο, παρέχω, πεζογράφος, πέψη, πίκρα, ποδιές, πράο, προοπτικές, προσαρμογή, προσοχής, προσπάθησα, προσφοράς, ράβδο, ρωμαϊκό, σεισμών, σκόνη, σκυτάλη, σπάμε, σπίθα, σπουδάσει, στάθμευαν, στέκω, σύγκριση, σύλληψη, συμβούν, συμμετέχει, συμπαθές, σύνταξη, σφαίρα, σωτήρ, τάγμα, ταχυδρόμος, τέντωμα, τετράδα, τίγρη, τόλμη, τραπεζάκι, τραπεζίτης, τριπλή, υδραυλικά, υδρόψυκτος, υπέγραψε, υπερβολικά, υπήκοος, ύποπτα, φέρε, φιλοδοξίες, φιλόξενες, φορολογικά, φραπέ, φταίμε, χαμογέλασα, χαρακτήρες, χρονική, χρυσοχοείο, ψήνονται, ψηφοθηρικό, ώσπου.

Pseudowords

αγκρισμός, αγλαρτός, αιμνέτη, αξήκας, αρτίμλικος, αυρακάνα, βαμυπέρα, βιλάνα, βοκένα, γαλτά, γαρονετίζω, γκαριμόκι, γκοιμένετε, γυτίσκα, δειξάπο, δεκλάμη, δήκωραν, δηλάκητα, διτρί, δογορονήσω, δουτρός, δυζίσεις, δώλκε, έγκα, ελδοπάδας, έλτικο, έντλος, ερνητευθεί, ερτωμίδα, εύτο, ευτρός, ζατέπα, ζενόπι, ζερελομίνα, ζομίθεται, ηγήτιβος, ηθέφαπο, ήκρο, ήμωρ, ηπρακάδι, ηρεξονίας, ησμήκας, θεβίρα, θέλμισες, ικόπαζα, κάδι, κατλής, κατσακόνα, κήξη, κλίπα, κλιτάρμι, κόρταμα, κούλτα, λαμήνιθος, λάργη, λέοις, λοθνί, λουτραμίδι, μαβόμεξη, μαγνί, μέκλη, μελδύο, μενθίς, μιργωτοί, μόπραθα, μπρόλι, νάβη, νεμίρω, νεύθας, ντρούσι, νυλτεπίνες, ξερύλτησα, οθλάνης, όλτι, ομόγεμπρο, ορεθός, οσμήκας, οστακούδα, παγκουλώνω, παδέφω, παζίστε, παζοβράφος, πακαÀζες, πακεινούδα, παλλίκωες, παράκυλθος, παριβάλα, πασάροψο, πατσακόνας, περιάμαιγη, περόγγαλος, πίρλα, πλοίμας, πολτέμνεος, πόντας, πόρμικο, πόσκεις, πούφτη, πρεσλάνι, πρευτά, προμάδι, προσγάτησα, προσμόβας, πυτρί, ράφθο, ρέλταμος, ρεφόρτι, ρονεîγω, σαίταρ, σάμες, σαντρέδες, σεκίπι, σικλί, σιρτεμάκι, σπουγάτει, στακώθητα, στάλμαυαν, σταμούδα, στεμακίδας, στογί, στοιλαμόκα, συλλαθή, σύμμαξους, συμπατέσει, συνθούν, ταβυτρόκος, τεγακόπα, τηλδί, τόμπρη, τούγι, τραμάκι, τριγκή, τρίδηκα, τσειμάνω, τσεκοτά, υγράπλικα, υμέγρασε, υπερμάφισα, ύπωντα, φανίζισε, φερτιμόγης, φιλογορίες, χανοβέκασα, χράτι, χρείπα, χρήτες, χρητήρας, ψησοκηρικό, ψόνα, ώλτου.

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix D

Results for the Linear Mixed Model Including RT Data From Both Tasks

		Random effects						
Group	Name	Variance	*SD*	Correlation				
iID	(Intercept)	6.70E–03	8.19E–02					
	Task	6.12E–03	7.83E–02	−.46				
sID	(Intercept)	3.52E–02	1.87E–01					
	Task	9.40E–02	3.07E–01	.31				
	iRT1	5.80E–03	7.62E–02	.50	.17			
	OrdS	1.98E–03	4.45E–02	.12	−.12	−.19		
	Task:iRT1	2.32E–02	1.52E–01	.26	.85	.23	−.23	
	Task:OrdS	6.30E–03	7.93E–02	.08	.11	−.15	.20	.05
Residual		5.20E–02	2.28E–01					

	Fixed effects				
Group	β	*SE*	*df*	*t*	*p*
(Intercept)	−1.35E+00	2.57E–02	247	−52.446	<.001
Task	−1.23E−01	3.77E–02	181	−3.270	.001
iRT1	2.66E−01	9.99E–03	134	26.648	<.001
iniPho1	−5.89E−02	1.47E–02	148	−4.000	<.001
iniPho2	7.23E−03	3.32E–02	148	.218	.828
iniPho3	3.44E−03	2.10E–02	149	.164	.870
iniPho4	2.07E−02	1.57E–02	149	1.319	.189
secPho1	−1.86E−02	1.84E–02	149	−1.011	.314
secPho2	2.74E−02	2.09E–02	149	1.311	.192
secPho3	1.74E−02	2.76E–02	150	.629	.530
secPho4	−2.33E−02	2.13E–02	148	−1.095	.275
secPho5	2.55E−02	1.96E–02	148	1.304	.194
OrdS	−2.15E−02	4.29E–03	127	−5.008	<.001
Education	−2.14E−02	6.69E–03	128	−3.193	.002
Nlet	2.13E−02	3.68E–03	149	5.802	<.001
sylfreqPho	4.35E−03	1.83E–03	149	2.383	.018
orthnei	−8.07E−03	4.79E–03	150	−1.686	.094
TaskOrder	−6.19E−02	1.48E–02	127	−4.191	<.001
logfreq	−3.10E−02	3.87E–03	149	−8.023	<.001
bigrSpe	1.25E−02	2.84E–02	149	.442	.659
gptrans	2.70E−04	9.54E–04	149	.283	.777
Task:iRT1	2.32E−01	2.00E–02	128	11.571	<.001
Task:iniPho1	−6.62E−02	1.48E–02	148	−4.473	<.001
Task:iniPho2	−1.89E−02	3.33E–02	148	−.568	.571
Task:iniPho3	−2.25E−02	2.11E–02	149	−1.066	.288
Task:iniPho4	5.65E−02	1.58E–02	148	3.586	<.001
Task:secPho1	2.75E−02	1.85E–02	148	1.487	.139
Task:secPho2	−2.01E−02	2.10E–02	148	−.957	.340
Task:secPho3	−1.33E−02	2.78E–02	151	−.479	.633
Task:secPho4	−3.66E−02	2.14E–02	148	−1.709	.090
Task:secPho5	3.34E−02	1.97E–02	148	1.699	.092
Education:Nlet	−2.11E−03	3.04E–04	35510	−6.934	<.001
Education:sylfreqPho	−1.27E−04	1.02E–04	35520	−1.246	.213
Education:orthnei	−5.39E−04	3.92E–04	35530	−1.377	.169
Task:Education	−1.04E−02	7.17E–03	126	−1.444	.151
Task:Nlet	8.99E−03	3.70E–03	148	2.432	.016
Task:sylfreqPho	−4.18E−03	1.84E–03	149	−2.274	.024
Task:TaskOrder	−3.76E−02	5.52E–02	128	−.681	.497
Task:logfreq	1.69E−02	3.87E–03	148	4.357	<.001
Task:orthnei	1.05E−02	4.76E–03	148	2.195	.030
Task:bigrSpe	−7.87E−03	2.85E–02	148	−.276	.783
Task:gptrans	1.29E−04	9.55E–04	148	.135	.893
TaskOrder:logfreq	6.03E−03	1.31E–03	35530	4.603	<.001

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix D (*continued*)

Group	Fixed effects				
	β	*SE*	*df*	*t*	*p*
Nlet:TaskOrder	−9.20E−03	1.27E–03	35520	−7.238	<.001
sylfreqPho:TaskOrder	−2.27E−03	4.26E–04	35510	−5.341	<.001
orthnei:TaskOrder	6.45E−03	1.66E–03	35520	3.897	<.001
TaskOrder:bigrSpe	−2.45E−02	9.37E–03	35540	−2.612	.009
TaskOrder:gptrans	−7.35E−04	3.11E–04	35530	−2.364	.018
orthnei:logfreq	9.59E−03	2.38E–03	149	4.033	<.001
Task:Education:Nlet	7.74E−04	6.08E–04	35510	1.273	.203
Task:Education:sylfreqPho	−3.95E−04	2.04E–04	35520	−1.940	.052
Task:TaskOrder:logfreq	−2.92E−03	2.60E–03	35540	−1.122	.262
Task:Nlet:TaskOrder	1.27E−03	2.54E–03	35520	.501	.617
Task:sylfreqPho:TaskOrder	4.19E−04	8.51E–04	35520	.493	.622
Task:orthnei:TaskOrder	–9.09E−03	3.25E–03	35530	−2.798	.005
Task:TaskOrder:bigrSpe	2.76E−02	1.88E–02	35540	1.470	.142
Task:TaskOrder:gptrans	2.70E−04	6.17E–04	35530	.437	.662
orthnei:TaskOrder:logfreq	−4.29E−03	9.20E–04	35520	−4.657	<.001

Note. Fit by maximum likelihood (BIC = −932.6, log likelihood = 902.1). The dependent variable is inverse reflected response time (iRT). Task coded as lexical decision = −.5, naming = +.5; iRT1 = (inverse reflected) RT to the preceding trial; OrdS = trial order (rescaled to 1/100th); iniPho = initial phoneme class (5 levels); secPho = second phoneme class (6 levels); TaskOrd = task order (coded as: first = −.5, second = +.5); logfreq = log frequency; Nlet = number of letters; orthnei = orthographic neighborhood size; sylfreqPho = phonological syllable frequency; gpcons = graphophonemic consistency; bigrSpe = letter bigram frequency; Education = years of formal education, excluding preschool/kindergarten.

Received February 12, 2015
Revision received July 3, 2015
Accepted July 27, 2015 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.