Word Length, Set Size, and Lexical Factors:
Re-examining What Causes the Word Length Effect

Dominic Guitard, Université de Moncton
Andrew J. Gabel, Memorial University of Newfoundland
Jean Saint-Aubin, Université de Moncton
Aimée M. Surprenant, Memorial University of Newfoundland
Ian Neath, Memorial University of Newfoundland

In press: Journal of Experimental Psychology: Learning, Memory, and Cognition
This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors’ permission. The final article will be available, upon publication, via its DOI: 10.1037/xlm0000551

Key words: word length effect, working memory, short-term memory, set size

Correspondence:

Ian Neath
Department of Psychology
Memorial University of Newfoundland
St. John’s, NL A1B 3X9
Canada
email: ineath@mun.ca
voice: 709 864 8495
fax: 709 864 2430
Abstract

The word length effect, better recall of lists of short (fewer syllables) than long (more syllables) words has been termed a benchmark effect of working memory. Despite this, experiments on the word length effect can yield quite different results depending on set size and stimulus properties. Seven experiments are reported that address these two issues. Experiment 1 replicated the finding of a preserved word length effect under concurrent articulation for large stimulus sets, which contrasts with the abolition of the word length effect by concurrent articulation for small stimulus sets. Experiment 2, however, demonstrated that when the short and long words are equated on more dimensions, concurrent articulation abolishes the word length effect for large stimulus sets. Experiment 3 shows a standard word length effect when output time is equated, but Experiments 4-6 show no word length effect when short and long words are equated on increasingly more dimensions that previous demonstrations have overlooked. Finally, Experiment 7 compared recall of a small and large neighborhood words that were equated on all the dimensions used in Experiment 6 (except for those directly related to neighborhood size) and a neighborhood size effect was still observed. We conclude that lexical factors, rather than word length per se, are better predictors of when the word length effect will occur.
Word Length, Set Size, and Lexical Factors:

Re-examining What Causes the Word Length Effect

When a list of words is recalled in order, the proportion of words recalled “is in an inverse relation to their syllabic length” (Calhoon, 1935, p. 620). This word length effect has been termed a “benchmark finding” that theories of short-term or working memory must address (Lewandowsky & Farrell, 2008), and indeed its importance arises, at least in part, because of its role as one of the four core phenomena that working memory was developed to explain (Baddeley, 1986). However, it has long been known that experiments on the word length effect can yield quite different results depending on the particular stimulus set used. In this paper, we report seven experiments that focus on set size and stimulus properties with the goal of addressing, at least in part, some discrepancies prominent in the literature. In Experiments 3-7, we wanted to use fairly large set sizes (up to 70 words per condition) to minimize any potential effect one odd word might have. If we use stimulus sets of only 6 or 7 words, a single unusual item in one condition could be sufficient to affect the result. However, there are a number of papers that suggest that the word length effect observed with large stimulus sets differs from the word length effect observed with small stimulus sets. Experiments 1 and 2 first address this issue.

Nairne (2002) noted the similarities among most accounts of the word length effect and described a “standard model” that embodies their common features. According to this model, items in short-term or working memory decay unless that decay is offset by subvocal rehearsal. Given the assumption that rehearsal speed is correlated with pronunciation time, more short items can be rehearsed than long items, which produces the basic word length effect. In the absence of rehearsal, recall of short and long words should be equivalent because the decay rate...
is assumed to be the same. This basic account can be found in many different models of working memory, including both simulation models (e.g., Page & Norris, 1998; Burgess & Hitch, 1999) and verbal models (e.g., Baddeley, 2000; Cowan et al., 1998).

Word Length and Stimulus Set

However, some word length effect results appear to depend critically on the particular stimulus set used. A first example of results specific to a particular stimulus set concerns an experiment reported by Baddeley, Thomson, and Buchanan (1975, Experiment IV). All theories based on the standard model predict a word length effect when the words are equated on all dimensions except for pronunciation time. Baddeley et al. created a set of 5 long and 5 short words that were equated for number of syllables, number of phonemes, and frequency, but which varied in pronunciation time. As predicted, the proportion of short words correctly recalled in order was higher than that for long words, 0.722 vs. 0.616. This result has been replicated many times using the same stimulus set (Cowan et al., 1992; Longoni, Richardson, & Aiello, 1993; Lovatt et al., 2000; Nairne, Neath, & Serra, 1997). However, all other published attempts to create another set of short and long words that differ only in pronunciation time find no difference in recall of short and long words (Caplan, Rochon, & Waters, 1992; Service 1998; Lovatt et al., 2000; Neath, Bireta, & Surprenant, 2003). Neath et al. concluded that there was likely some idiosyncratic property of one word in the original set which, given the small number of items, may have been sufficient to cause a difference in recall.

A second example of results specific to particular stimulus sets examines recall of lists that contain both short and long words. Cowan, Baddeley, Elliot, and Norris (2003) reported that mixed lists of both long and short words were recalled worse than pure short lists but better than
pure long lists. In contrast, Hulme, Surprenant, Bireta, Stuart, and Neath (2004) found that mixed lists were recalled equally as well as pure short lists. Bireta, Neath, and Surprenant (2006) demonstrated that the different results were due to the stimulus sets used: They replicated the Cowan et al. results when using the Cowan et al. stimuli, and replicated the Hulme et al. results when using the Hulme et al. stimuli.

Word Length and Set Size

A third example has been presented in the literature as an example of a different word length effects as a function of the size of the stimulus set but, as we argue below, we think the different results are more likely due to stimulus set properties. All accounts based on the standard model predict that when subvocal rehearsal is prevented, the word length effect should be attenuated or abolished. Abolition is predicted only if the rehearsal is completely prevented; if rehearsal is reduced but not completely eliminated, then a small advantage may still accrue to short words. The key prediction, then is observing an interaction, although most studies that have examined this issue focus on either preservation or abolition of the word length effect.

Baddeley et al. (1975, Experiment VII) reported an experiment that demonstrated a word length effect when using a pool of 10 short (one-syllable) and 10 long (five-syllable) words. However, when subjects were asked to perform concurrent articulation 1 – repeatedly saying some irrelevant items, such as the numbers 1-4, out loud over and over – during list presentation, the word length effect was abolished. This result has been replicated many times when similarly small stimulus sets are used (Baddeley, Lewis, & Vallar, 1984; Bireta, Fine, & VanWormer, 1986).

1 We prefer the term “concurrent articulation” to the more common “articulatory suppression” because it describes the manipulation rather than the assumed consequence.
2013; Longoni, Richardson, & Aiello, 1993). LaPointe and Engle (1990, Experiments 4 and 5) replicated this abolition of the word length effect when they used a small pool of 8 short and 8 long words, but found a preserved word length effect despite concurrent articulation when they used a large pool of 81 short and 81 large word. Russo and Grammatopoulou (2003, Experiment 5) replicated this preserved word length effect when using a large pool of short (1 syllable) and long (4 or 5 syllables) words. Bhatarah, Ward, Smith, and Hayes (2009, Experiment 3) used lists of 8 short (1 syllable) or 8 long (5 syllable) words drawn from a large pool and found a significant interaction between word length and concurrent articulation, but still found a preserved word length effect.2

The interaction between stimulus pool size and concurrent articulation on word length studies is particularly striking because the size of the stimulus pool, large or small, has no such effect on other manipulations that interact with concurrent articulation. For example, the acoustic confusion effect (Conrad, 1964) refers to the finding that lists made up of dissimilar-sounding items (e.g., FKXMQ) are better recalled than lists of similar-sounding items (e.g., BPTDVC). Just as with the word length effect, concurrent articulation eliminates this effect when the to-be-remembered items are presented visually (Murray, 1968). Whereas the early studies used small stimulus sets (typically letters), later studies used large sets and obtained the same result: concurrent articulation eliminated the acoustic similarity effect (V. Coltheart, 1993; Surprenant, Neath, & LeCompte, 1999).

Because the standard model cannot explain the preserved word length effect under

---

2 Romani, McAlpine, Olson, Tsouknida, and Martin (2005, Experiment 1) reported the abolition of the word length effect with an open pool but had concurrent articulation during both presentation and test.
concurrent articulation when a large stimulus pool is used, theorists posited two different word length effects and two different explanations. The standard model was still used to explain word length effects with small pools but new accounts were offered for word length effects with large pools. For example, LaPointe and Engle (1990) suggested that whereas phonological codes are used with small pools, different codes are used with large pools. As a second example, Cowan et al. (1998) suggested in addition to its other uses, the rehearsal process might also be used to link items in order. When concurrent articulation prevents the use of the rehearsal process, the phonological representations are no longer sufficient to support ordered recall. Thus, concurrent articulation eliminates the word length effect for small pools. However, with a large pool, the phonological representations "might often be able to move from an inactive state to an active state and be linked to other items in the list, even in the presence of articulatory suppression" (p. 155). If this occurs, concurrent articulation will not be able to eliminate the word length effect.

As a third example, V. Coltheart (1993) suggested that concurrent articulation minimizes both phonological encoding and rehearsal, and therefore recall is largely based on orthographic and unelaborated semantic codes rather than phonological codes. It is assumed that such codes are more difficult to establish for long than for short words, giving an advantage to short words in the no articulation condition. With small pools, it is more likely that orthographic and/or semantic codes can be built up over time because the word is repeated, resulting in performance based on different types of encoding compared to when large pools are used. As a final example, Russo and Grammatopoulou (2003) discussed a theoretical account that shares similarities with the above.

The need for two different explanations for small and large pools depends on whether there really are two different word length effects, one of which interacts with concurrent
articulation and one of which does not. The literature is equivocal on this point: Although three studies using large pools found a preserved word length and numerous studies with small pools found an abolished word length effect, it is not clear that set size is the key variable: It is possible that differences in the specific stimulus sets used give rise to the difference.

Unfortunately, only one of the studies that found a preserved word length effect under concurrent articulation with a large stimulus pool provided details of the words used. Inspection reveals that the LaPointe and Engle (1990) stimuli differ not only in length but also in word frequency, with the short words being more frequent. The stimuli are shown in the appendix. It is well established that high frequency words are better recalled on immediate serial recall tasks than low frequency words (Poirier & Saint-Aubin, 1996; Roodenrys, Hulme, Lethbridge, Hinton, & Nimmo, 2002; Watkins & Watkins, 1977). Thus it is entirely possible that set size is not the causal factor, but rather, that the prior demonstrations of differences arose because the short and long words differed on dimensions other than length. That is, we speculate that the word length effect remained under concurrent articulation because the short words were of higher frequency than the long words.

The purpose of the first two experiments, then, is to test this hypothesis. The prediction is that Experiment 1, which uses the same large stimulus set used by LaPointe and Engle (1990), will replicate their finding of a word length effect in both the control and concurrent articulation conditions. Experiment 2 is identical, except that the short and long words are equated for frequency; the prediction now is that the word length effect will be abolished by concurrent articulation, despite the fact that a large set size is being used.

One reason why it is necessary to make sure that the word length effect is abolished by

---

3 We discuss differences on other dimensions shown in the appendix later in the paper.
concurrent articulation for large stimulus pools is because we want to use large stimulus pools to examine other stimulus properties. With only 5 or 6 or 8 words, it takes only 1 unusual word to potentially give rise to a mnemonic difference that may have nothing to do with length. In contrast, as the set of words gets larger, the effects of one odd word will be greatly minimized. In particular, we wish to investigate various lexical factors. We now set out the reason why we think that lexical factors, including orthographic neighborhood size, might be better predictors of when a word length effect will be observed than length itself.

Word Length Effect and Orthographic Neighborhood Size

An orthographic neighbor of a target word is a valid word that differs by a single letter and the set of these words is referred to as the target word’s neighborhood (M. Coltheart, Davelaar, Jonasson, & Besner, 1977). For example, given the target word *dog*, orthographic neighbors include *fog*, *dig*, *dot*, and so on. A number of studies have shown that lists comprised of words with larger orthographic neighborhoods are better recalled than lists comprised of words with smaller orthographic neighborhoods (Allen & Hulme, 2006; Derraugh, Neath, Surprenant, Beaudry, & Saint-Aubin, 2017; Glanc & Greene, 2012; Roodenrys et al., 2002).

One explanation of the orthographic neighborhood size effect focuses on a facilitative effect at the redintegration stage (Derraugh et al., 2017; Roodenrys, 2009). A degraded representation of a word partially activates or primes its neighbors. In the case of *fir*, for example, there are 11 neighbors, each of which share two letters. The partially-activated neighbors produce feedback that strengthens *fir* because they differ by only 1 letter. In contrast, in the case of *elm*, there are only 3 neighbors, so its degraded representation receives far less feedback and is consequently not nearly as strengthened as *fir*. 
A problem arises because short words tend to have more neighbors than long words. Jalbert, Neath, Bireta, and Surprenant (2011) noted that this confounding of length and neighborhood size is present in all published word length studies that report the stimuli, regardless of the stimulus set size. Looking at previous stimulus sets, Jalbert et al. calculated that the mean orthographic neighborhood size of short words in the existing literature was 8.61 compared to 0.24 for long words. Given this confound, it is not possible to state whether word length effect papers are observing an effect of word length or an effect of neighborhood size or, as we discuss further below, an effect of other lexical factor (or combination of factors) that also varies with length.

Jalbert et al. (2011a) reported two experiments that showed that when two different sets of short (1 syllable) and long (3 syllable) words were equated for orthographic neighborhood size (as well as concreteness, familiarity, imageability, frequency, and frequency of the orthographic neighbors), the word length effect disappeared. Jalbert, Neath, and Surprenant (2011) reported an experiment showing that the neighborhood size effect is eliminated by concurrent articulation. Both of these results suggest that neighborhood size could be a plausible explanation of the word length effect. However, the studies in the two Jalbert et al. papers used only a small pool of items, which increases the possibility that one or two odd words were driving the results. One purpose of the remaining experiments is to replicate these results using larger stimulus pools.

---

4 Their study also included a manipulation of whether the list was pure (i.e., only large neighborhood or only small neighborhood words) or mixed (i.e., a combination of large and small neighborhood words within the same list). For the pure lists, the interaction between neighborhood size and concurrent articulation was significant, $F(1,30) = 7.919, MSE = 0.003, \eta^2_p = 0.209, p < 0.01$, although this particular analysis was not reported in the paper.
In addition to the issue of set size, a second potential objection to the conclusions of Jalbert et al. (2011a,b) is that they may not have equated the words on all the appropriate dimensions. Storkel (2004) noted that a large number of lexical measures are correlated with length. Some measures are negatively correlated with length, such as number of orthographic neighbors, whereas others are positively correlated with length, such as constrained and unconstrained unigram, bigram, and trigram counts. A constrained unigram is a specific letter in a specific position in a word of a specific length. For example, the o in stoat is considered the same as the o in float (same position, same length word) but is not considered the same as the o in coach (same length word but different position) or the o in violin (same position but different length). An unconstrained unigram allows the letter to be in any position of a word of the same length, so the o in coach is an unconstrained unigram of stoat. The same definitions hold for bigrams and trigrams except these count 2 or 3 letters, respectively. As an example of a constrained trigram, the oat in stoat is considered the same as the oat in float whereas it is not considered to be the same as the oat in boat. The latter, however, is an unconstrained trigram.

Storkel (2004) examined three different ways of taking these correlations into account and recommended using z scores. To compute the relevant z score, one first gets the mean and standard deviation of the measure for words at each length of interest. Then, for each word, a z score is computed by calculating the difference between the value of the target word and the mean and then dividing by the standard deviation. In the appendix, we report 7 such measures: one that corresponds to orthographic neighborhood size (OrthZ); one each that corresponds to constrained unigram, bigram, and trigram frequencies (C1Z, C2Z, and C3Z, respectively); and one each that corresponds to unconstrained unigram, bigram, and trigram frequencies (U1Z, U2Z, and U3Z, respectively). It turns out that Jalbert et al. did not equate the short and long
words for the $z$-transformed measure of number of orthographic neighbors. For example, in Experiment 3 of Jalbert et al. (2011a), the mean OrthZ score for short words was -0.795 compared to +0.507 for the long words. The word length effect may have been eliminated because the long words were higher on this measure than the short words. The purpose of Experiments 4 and 5 was to demonstrate that even when a set of words that vary in length (2 vs 3 syllable words in Experiment 4; 1 vs 3 syllable words in Experiment 5) are equated for these $z$-score based measures, no word length effect emerges.

A final problem is that there may be issues concerning measures based on Coltheart’s $N$, which is the measure of orthographic neighborhood size described above, and is a simple count of the number of words of the same length that differ by one letter. Yarkoni, Balota, and Yap (2008) note that $N$ has some possible limitations. First, it is a binary measure, in that a word either is or is not a neighbor. Their proposed alternative allows a word to be “more or less neighborly” (p. 971) because it is based on Levenshtein distance, a measure of how many edits must be performed to change one string into another string (Levenshtein, 1966). Thus, their measure takes into account how far from the target the potential neighbor might be by including this distance measure. Second, $N$ includes only those words that are of the same length.\footnote{Other measures include the addition or deletion of a letter (e.g., Luce & Pisoni, 1998), but they are still binary measures.} Yarkoni et al. give as examples the close relationship (in terms of priming) of widow and window, and of trail and trial, the latter an example of two words that are not neighbors according to Coltheart’s $N$ because of a reversal of letters. Yarkoni et al.’s measure is based on the minimum number of edits – either addition, deletion, insertion, or substitution – needed to change one word into another. Specifically, they computed the mean Levenshtein distance for the target word’s 20
closest neighbors. They limited the measure to 20 words because almost no additional predictive ability was gained by including more words. An analysis found that this measure, OLD20, was a better predictor of a number of findings, such as lexical decision RTs, than \( N \). A companion measure, OLDF, indicates the frequency of the 20 words. Experiment 6 compared recall of short (1 syllable) and long (3 syllable) words when equated on a total of 18 different measures, including OLD20 and OLDF. Finally, Experiment 7 compared recall of a small and large neighborhood words that were equated for length and also for all the dimensions used in Experiment 6 (except for those directly related to neighborhood size) to determine whether a neighborhood size effect would be observed.

**Experiment 1**

The purpose of Experiment 1 was to replicate, using immediate written strict serial recall, the LaPointe and Engle (1990, Experiment 5) finding of a preserved word length effect under concurrent articulation when using a large pool of words and also to assess whether there was a significant interaction between word length and articulation condition because LaPointe and Engle did not perform that analysis. This and all subsequent experiments were approved by Memorial University’s Interdisciplinary Committee on Ethics in Human Research.

**Method**

**Subjects**

Twenty-five undergraduates from Memorial University of Newfoundland participated in exchange for an honorarium. The mean age was 20.64 (\( SD = 3.56 \), range 18-34), 18 identified themselves as female and 7 as male, and all identified themselves as native speakers of English.

**Stimuli**

The stimuli were the 81 short (one syllable) and 81 long (two to five syllable) words from
LaPointe and Engle (1990, Exp. 5). In this experiment, sampling from each set was performed without replacement such that a given word appeared only once during the entire experiment. The stimuli are shown in the appendix.

Procedure

The subjects used a mouse to click on a “Start next trial” button on a computer screen. One second after the fixation point disappeared, a message was shown for 3 s. The message was either “Please start saying 1 2 3 4 1 2 3 4 over and over” or “Please do not say anything out loud.”

One second after the message disappeared, 6 words were shown one at a time for 1 s each in uppercase in the center of the screen. After the final word had been shown, subjects were informed they could stop the articulation task (if it were being performed) and were asked to write down the words on their answer sheet in strict serial order. They were informed that they needed to write down the first word first, the second word second, and so on. They were further informed that they could not go back and change an answer. If they were unsure of a response they were encouraged to guess or else leave that slot on the answer sheet blank. The subjects initiated the next trial when ready.

There were 2 unscored practice trials. One trial had concurrent articulation instructions and the other had keep silent instructions. In addition, one of these trials used short words and one used long words. The particular combination of stimulus type and secondary task instruction on the practice trials was determined randomly for each subject. Of the 20 scored trials, half had concurrent articulation instructions and half had keep silent instructions. There were equal numbers of short and long trials within each of these. The order of the conditions was randomly determined for each subject, and the words used on each trial was also randomly determined for
each subject. An experimenter remained in the room to ensure compliance with the instructions.

Results

The proportion of words correctly recalled in order in the quiet condition and the concurrent articulation condition are shown in the left and right panels of Figure 1, respectively. As can be seen, short words were better recalled than long words in both conditions, a conclusion supported by the following analyses.

The data were analyzed using a two word lengths (short vs. long) × two articulation conditions (no articulation vs. concurrent articulation) × six serial positions repeated measures analysis of variance. There was a significant main effect of word length, $F(1,24) = 28.083, MSE = 0.049, p < 0.01$. The mean proportion of short words correctly recalled in order was 0.429 ($SD = 0.120$) compared to 0.333 ($SD = 0.103$) for the long words, Cohen’s $d = 0.852$. There was also a significant main effect of articulation condition, $F(1,24) = 59.338, MSE = 0.087, p < 0.01$. The mean proportion of words correctly recalled in order in the quiet condition was 0.474 ($SD = 0.134$) compared to 0.288 ($SD = 0.102$) in the concurrent articulation condition; $d = 1.564$. There was also a significant main effect of serial position, $F(5,120) = 49.756, MSE = 0.072, p < 0.001$, as is typical of immediate serial recall.

Of particular importance was the significant interaction between word length and articulation condition, $F(1,24) = 4.654, MSE = 0.043, p < 0.01$. In the quiet condition, the difference between recall of the short and long words was $0.540 – 0.408 = 0.132$, $d = 0.897$. In the concurrent articulation condition, the difference was $0.317 – 0.259 = 0.058$, $d = 0.490$. Following LaPointe and Engle (1990), we used planned $t$-tests to assess whether the difference between short and long words was significant in each articulation condition. In both cases the difference was significant; no articulation, $t(24) = 5.448, p < 0.01$, and concurrent articulation,
The interaction between articulation condition and serial position was significant, $F(5,120) = 9.628$, $MSE = 0.050$, $p < 0.001$, as can be seen in Figure 1. Neither the word length × serial position nor the three-way interactions were significant, $Fs < 1.0$.

Discussion

LaPointe and Engle (1990) reported a significant word length effect in a span task when the to-be-remembered items came from a large pool regardless of whether the subjects engaged in concurrent articulation. Experiment 1 replicated this result using the same stimuli but using a method more frequently used to study working memory, immediate strict serial recall. In addition to replicating the results, Experiment 1 also found that the interaction between word length and articulation condition was significant. Although concurrent articulation did not eliminate the word length effect, it did reduce the size of the effect.

The means observed in Experiment 1 are similar to those observed by LaPointe and Engle: In their study, the magnitude of the word length effect was $0.57 - 0.42 = 0.15$ in the no articulation condition compared to $0.30 - 0.23 = 0.07$ in the concurrent articulation condition. This suggests that concurrent articulation did indeed reduce the magnitude of the word length effect in their study. However, because they used planned contrasts instead of an ANOVA, the interaction between word length and concurrent articulation was not reported. If one assumes that there was an interaction between word length and concurrent articulation in their study, the pattern of results in the literature becomes more clear. One study, Russo and Grammatopoulou (2003, Experiment 5) did not observe an interaction, but three studies – Bhatarah et al. (2009), LaPointe and Engle (1990), and Experiment 1 in this paper – did find a significant interaction. It now becomes reasonable to conclude that concurrent articulation interacts with word length for
both small and large pools. Note that this result is readily accommodated by the standard model: if rehearsal is reduced but not completely eliminated, then a small advantage may still accrue to short words. It is only when rehearsal is completely prevented that abolition is predicted.

Although we found a significant interaction in Experiment 1, the word length effect remained under concurrent articulation. Our hypothesis is that the word length remained both in this experiment and in that of LaPointe and Engle (1990) because the short words were of higher frequency than the long words. If this hypothesis is correct, it follows that if a large pool is used in which the short and long items are equated for frequency, then in addition to a significant word length × articulation condition interaction, the word length effect should be abolished in the concurrent articulation condition.

Experiment 2

Experiment 2 was identical to Experiment 1 except that a new set of stimuli was used in which the short and long words were equated for frequency.

Subjects

Twenty-five different undergraduates from Memorial University of Newfoundland participated in exchange for an honorarium. The mean age was 20.32 (SD = 3.54, range 18-34), 18 identified themselves as female and 7 as male, and all identified themselves as native speakers of English.

Stimuli

The stimuli were 101 one-syllable and 101 two and three-syllable nouns drawn originally from a much larger pool sampled from M. Coltheart (1981). The original pool was reduced in size until the short and large words were equated for concreteness, imageability, familiarity (from Coltheart); CELEX frequency (from Medler & Binder, 2005); and SUBTLEXUS frequency
and contextual diversity (from Brysbaert & New, 2009). The short and long words differed significantly in length as measured by the number of letters, the number of phonemes, and the number of syllables. They also differed in a number of lexical measures, as did the words in Experiment 1. Details, including statistics comparing the short and long words, are provided in the appendix.

Procedure

The procedure was identical to that of Experiment 1.

Results

The proportion of words correctly recalled in order in the quiet condition and the concurrent articulation condition are shown in the left and right panels of Figure 2, respectively. As can be seen, short words were better recalled than long words only in the quiet condition, a conclusion supported by the following analyses.

The data were analyzed using a two word lengths (short vs. long) × two articulation conditions (no articulation vs. concurrent articulation) × six serial positions repeated measures analysis of variance. There was a significant main effect of word length, $F(1,24) = 19.667, MSE = 0.035, p < 0.01$. The mean proportion of short words correctly recalled in order was 0.413 ($SD = 0.114$) compared to 0.345 ($SD = 0.102$) for the long words, $d = 0.627$. There was also a significant main effect of articulation condition, $F(1,24) = 44.346, MSE = 0.085, p < 0.01$. The mean proportion of words correctly recalled in order in the quiet condition was 0.459 ($SD = 0.125$) compared to 0.300 ($SD = 0.110$) for the concurrent articulation condition, $d = 1.349$. There was also a significant main effect of serial position, $F(5,120) = 35.046, MSE = 0.064, p < 0.001$, reflecting a serial position function typical of immediate serial recall.

Of particular importance was the significant interaction between word length and
articulation condition, $F(1,24) = 9.880, MSE = 0.060, p < 0.01$. In the quiet condition, the difference between recall of the short and long words was $0.524 – 0.393 = 0.131, d = 0.919$. In the concurrent articulation condition, the difference was $0.303 – 0.297 = 0.006, d = 0.043$.

Following LaPointe and Engle (1990) and as in Experiment 1, we used planned $t$-tests to assess whether the difference between short and long words was significant in each articulation condition. In the quiet condition, the difference was significant, $t(24) = 4.802, p < 0.01$, but in the concurrent articulation, the difference was not significant, $t(24) = 0.233, p > 0.45$.

The interaction between articulation condition and serial position was again significant, $F(5,120) = 9.880, MSE = 0.060, p < 0.001$, as can be seen in Figure 2. The word length $\times$ serial position did not quite reach the adopted significance level, $F(5,120) = 1.861, MSE = 0.042, p = 0.072$. The three-way interaction was not significant, $F < 1.0$.

Discussion

In Experiment 1, there was a significant interaction between word length and articulation condition even though the word length effect remained under concurrent articulation. However, the short words were of higher frequency than the long words. In Experiment 2, the short and long words were equated for frequency and the word length effect was eliminated by concurrent articulation. This finding is consistent with the idea that the preservation of the word length effect observed by LaPointe and Engle (1990) and in Experiment 1 was due to the difference in word frequency between the short and long words. On this point, it worth nothing that the word frequency effect has been found to be immune from concurrent articulation with the frequency effect as large under concurrent articulation as in the control condition (Tehan & Humphreys, 1988; Neath, Surprenant, Gabel, & Seffinga, 2016).

If one assumes that the interaction between word length and articulation condition in
LaPointe and Engle (1990) would have been significant if assessed, then Experiment 2 means there are now 4 demonstrations that concurrent articulation interacts with word length when large stimulus sets are used in the same way as when small stimulus sets are used. Moreover, this pattern of interactions also parallel those observed when acoustic similarity is manipulated rather than word length (V. Coltheart, 1993; Surprenant et al., 1999). Given these results, we argue that there is no need for different theoretical explanations of word length effects as a function of set size.

It is important to note that we are not claiming that set size has no effect on memory; rather, we are claiming that articulation condition interacts with word length in the same way for both small and large pools. It is also important to note that the significant interactions include both a preserved word length effect (e.g., LaPointe & Engle, 1990, Experiment 5; Experiment 1; Bhatarh et al., 2009) and an abolished word length effect (Experiment 2). The only data that do not fit this pattern are those of Russo and Grammatopoulou (2003), who did not find an interaction; in the absence of other evidence, we speculate that it was some property of their stimulus set that produced the divergent findings.

Together, Experiments 1 and 2 provide evidence against the view that word length effects differ as a function of set size. This is important because it allows us to use large set sizes, and using large set sizes is important because it minimizes the effects of any one or two odd or unusual words. However, as noted earlier and as can be seen in the appendix, even though the stimuli used in the first two experiments were equated on a large number of dimensions, the stimuli did vary significantly on other dimensions in addition to the critical dimension of interest, word length. One reason for the differences is that we were trying to emulate previous studies, and in those studies such confounds are evident. The older a study, the more likely it is that the
researchers were unaware of the importance of some dimension at the time the work was conducted. A second reason is that as the number of dimensions under consideration increases, so too does the difficulty in equating the stimuli on all of them. Before spending the time and effort to equate the two sets of stimuli on as many dimensions as we could, we thought it prudent to clarify whether set size affects the pattern of results.

The purpose of the remaining studies, then, is to increasingly better equate the short and long words on more dimensions. The first major change was to switch from immediate written strict serial recall to immediate strict serial reconstruction of order in order to control for output time. The former test confounds output time with word length because it takes longer to write down long words than short words. In a reconstruction of order test, it takes the same amount of time to click on a button with a short label as one with a long label. Moreover, it allows us to measure output time and confirm that the words are equated for this dimension. The second major change was to control for differences in lexical factors using the measures suggested by Storkel (2004).

Before reporting those experiments, for which the prediction is no difference in recall of short and long words, we needed to confirm that standard word length effects (using confounded stimuli) are observable with our procedure. Experiment 3, then, compared recall of 2 versus 3 syllable words that remained confounded on a number of lexical dimensions, just like the stimuli in Experiments 1 and 2, and the prediction was that a word length effect would obtain. Experiment 4 was similar to Experiment 3 except that the short and long words were now equated for not only Coltheart’s N but also for the z-transformed orthographic measures. Experiment 5 was identical to Experiment 4 but compared recall of 1 versus 3 syllable words rather than 2 versus 3 syllable words. Experiment 6 was identical to Experiments 4 and 5 but
WORD LENGTH AND NEIGHBORHOOD SIZE

compared recall of 2 and 4 syllable words that were equated on OLD20, the non-binary measure of orthographic neighborhood size suggested by Yarkoni et al. (2007). Finally, Experiment 7 used words that were equated on 18 different dimensions, including length, but which varied only on neighborhood size to ensure that the neighborhood size effect is observable with such highly controlled stimuli.

Experiment 3

Experiments 1 and 2 used immediate written strict serial recall because the focus was on demonstrating that large pools of long and short words interact with articulation condition in the same way as small pools of long and short words, and in the literature, written recall is most often used. One problem with both written and spoken recall, however, is that longer words take longer to output than shorter words. In Experiment 3 the task was changed to immediate strict serial reconstruction of order so that output time could be measured and reported. The list length was also increased from 6 to 7 as a result of the change in test. As in Experiments 1 and 2, the stimuli were designed to reflect the typical word length effect stimuli in that the words varied in both length (in this case, 2 versus 3 syllables) as well as a number of other lexical dimensions. The prediction from both the standard model and the lexical factors hypothesis is that short words will be better recalled than long words, despite controlling for output time.

Subjects

Twenty-five volunteers from ProlificAC were paid £10.00 per hour. For this and all subsequent experiments reported here, the following inclusion criteria were used: (1) Native speaker of English; (2) approval rating of at least 90% on prior submissions at ProlificAC; and (3) age between 19 and 39. The mean age was 27.20 ($SD = 4.56$, range 20-34) and 16 identified
themselves as male, 5 identified themselves as female, and 4 did not include a response to that question.

**Stimuli**

The stimuli were 36 two-syllable words and 36 three-syllable words. The words were drawn originally from a much larger pool sampled from M. Coltheart (1981), and the original pool was reduced in size until the large and small neighborhood words were equated for concreteness, imageability, familiarity (from Coltheart); CELEX frequency (from Medler & Binder, 2005); and SUBTLEX US frequency and contextual diversity (from Brysbaert & New, 2009). The short and long words did differ on many lexical dimensions, although not as many as in Experiments 1 and 2. Details, including statistics comparing the large and small neighborhood words, are provided in the appendix.

**Procedure**

The subjects used a mouse to click on a “Start next trial” button on a computer screen. One second after the fixation point disappeared, 7 words were shown one at a time for 1 s each in uppercase in the center of the screen. After the final word had been shown, seven buttons were labelled, in alphabetical order, with the words just seen. The task was to click on the buttons to recreate the presentation order. Once a button had been clicked, its icon changed from an active to an inactive state and could not be selected again. Once seven responses had been made, the “Start next trial” button became active. Subjects could take self-paced breaks by waiting to click on this button.

There were 20 lists, half with short words and half with long words. The order of the lists was randomly determined for each subject. For each list, 7 of the 36 words were randomly selected from the appropriate pool. Thus, unlike Experiments 1 and 2, a given word could appear
in multiple lists, although never more than once in a given list, and on average, a given word appeared approximately twice during the experiment.

Results and Discussion

The proportion of short and long words correctly placed in order (left panel) and the mean output time (right panel) are shown in Figure 3. As can be seen, there was a word length effect, with better recall of short compared to long words, even though output time was equivalent. These observations were confirmed by the following analyses.

The data were analyzed using a two word length (short vs. long) × seven serial positions repeated measures analysis of variance. There was a significant main effect of word length, $F(1,24) = 12.054, \text{MSE} = 0.039, p < 0.01$. The mean proportion of short words correctly recalled in order was 0.581 ($SD = 0.201$) compared to 0.507 ($SD = 0.213$) for the long words, $d = 0.356$. There was also a significant main effect of position, $F(6,144) = 22.803, \text{MSE} = 0.028, p < 0.01$. The interaction was not significant, $F(6,144) < 1$.

The mean output time data were also analyzed using a two word length (short vs. long) × seven serial positions repeated measures analysis of variance. The main effect of length was not significant, $F(1,24) < 1$. The mean output time for short words was 1689.5 ms ($SD = 404.7$) compared to 1710.4 ms ($SD = 445.7$) for long words, $d = 0.049$. There was a significant main effect of position, $F(6,144) = 52.165, \text{MSE} = 788108.04, p < 0.01$. The interaction was not significant, $F(6,144) = 1.308, \text{MSE} = 134004.09, p > 0.25$.

Experiment 3 demonstrated that a word length effect is observable with the current procedure, even though output time was equated. This is important because it rules out a number of potential objections to the lack of a word length effect in the next three experiments. However, as in Experiments 1 and 2, word length was confounded with a number of lexical factors. In the
Experiment 4

Experiment 3 demonstrated that a typical word length effect obtains when people are asked to indicate the order of 2- and 3-syllable words. In Experiment 4, we again compared recall of lists of 2 versus 3 syllable nouns that were equated for concreteness, familiarity, imageability, frequency, and contextual diversity, and were also equated for output time. Unlike in Experiment 3, though, the short and long words in Experiment 4 were also equated for orthographic neighborhood, the $z$-transformed measure of orthographic neighborhood size, and also the $z$-transformed measures of constrained and unconstrained unigram, bigram, and trigram counts. The only measures, other than length (number of letters, number of phonemes, and number of syllables) that differed between the short and long words were the Levenshtein distance based measures of neighborhood size and frequency (OLD20 and OLDF). The prediction of the standard model is that a word length effect should obtain whereas the prediction from the lexical factors hypothesis is that performance should be equivalent.

Subjects

Twenty-five different volunteers from ProlificAC, with the same inclusion criteria as Experiment 3, were paid £10.00 per hour. The mean age was 29.68 ($SD = 5.03$, range 20-38) and 15 identified themselves as female, and 9 identified themselves as male.

Stimuli

The stimuli were 36 two-syllable words and 36 three-syllable words. The words were drawn originally from a much larger pool sampled from M. Coltheart (1981), and the original pool was reduced in size until the large and small neighborhood words were equated for
concreteness, imageability, familiarity (from Coltheart); CELEX frequency, orthographic neighborhood size, and frequency of orthographic neighbors (from Medler & Binder, 2005); and SUBTLEX\textsubscript{US} frequency and contextual diversity (from Brysbaert & New, 2009). In addition, the short and long words were equated on various z-transformed orthographic measures; details, including statistics comparing the short and long words, are provided in the appendix.

**Procedure**

The procedure was the same as in Experiment 3.

**Results and Discussion**

The proportion of short and long words correctly placed in order (left panel) and the mean output time (right panel) are shown in Figure 4. As can be seen, there was no word length effect, with recall of short and long words and output times equivalent. These observations were supported by the following analyses.

The data were analyzed using a two word length (short vs. long) × seven serial positions repeated measures analysis of variance. The main effect of word length was not significant, $F(1,24) < 1$. The mean proportion of short words correctly recalled in order was 0.491 ($SD = 0.169$) compared to 0.492 ($SD = 0.177$) for the long words, $d = 0.003$. There was a significant main effect of position, $F(6,144) = 35.863$, $MSE = 0.033$, $p < 0.01$. The interaction was not significant, $F(6,144) < 1$.

The mean output time data were also analyzed using a two word length (short vs. long) × seven serial positions repeated measures analysis of variance. The main effect of length was not significant, $F(1,24) < 1$. The mean output time for short words was 1615.7 ms ($SD = 752.6$) compared to 1625.3 ms ($SD = 762.8$) for long words, $d = 0.013$. There was a significant main effect of position, $F(6,144) = 42.535$, $MSE = 628297.97$, $p < 0.01$. The interaction was not
significant, $F(6,144) < 1$.

Experiment 3 demonstrated a word length effect with better recall of 2 compared to 3 syllable words. Using the same methods and procedure, Experiment 4 demonstrated that when 2 and 3 syllable words were equated for a number of lexical factors, the word length effect disappeared.

It is important to note that the stimuli in the two experiments differed in a number of ways. In Experiment 3, the short and long words were not equated for the number of orthographic neighbors (Coltheart’s $N$), frequency of the neighbors, the $z$-transformed constrained trigram count, and the $z$-transformed unconstrained trigram count, whereas in Experiment 4, the short and long words were equated on all of these dimensions. Therefore, it is not logically possible to isolate one of these dimensions, or some combination, as the causative factor. Nonetheless, the important result is the absence of a word length effect. Although it is difficult to argue that a difference of only 1 syllable is insufficient to produce a word length effect given the results of Experiment 3, we nevertheless designed Experiment 5 as a replication of Experiment 4 but the difference between the short and long words was increased to 2 syllables.

**Experiment 5**

Experiment 5 was almost identical to Experiment 4 except that the short words differed by 2 syllables rather than just 1. The short and long words were again equated for concreteness, familiarity, imageability, frequency, contextual diversity, output time, orthographic neighborhood, the $z$-transformed measure of orthographic neighborhood size, the frequency of the neighbors, as well as the $z$-transformed measures of constrained and unconstrained unigram, bigram, and trigram counts. As in Experiment 4, the only measures, other than length (number of
letters, number of phonemes, and number of syllables) that varied were OLD20 and OLDF. The prediction of the standard model is that a word length effect should obtain whereas the prediction from the lexical factors hypothesis is that performance should be equivalent.

Subjects

Twenty-five different volunteers from ProlificAC, with the same inclusion criteria as Experiment 4, were paid £10.00 per hour. The mean age was 29.48 (SD = 5.55, range 21-38) and 17 identified themselves as female, and 3 identified themselves as male; 5 did not indicate a response to that question.

Stimuli

The stimuli were 30 one-syllable words and 30 three-syllable words. The words were drawn originally from a much larger pool sampled from M. Coltheart (1981), and the original pool was reduced in size until the large and small neighborhood words were equated for concreteness, imageability, familiarity (from Coltheart); CELEX frequency, orthographic neighborhood size, and frequency of orthographic neighbors (from Medler & Binder, 2005); and SUBTLEXUS frequency and contextual diversity (from Brysbaert & New, 2009). In addition, the short and long words were equated on various z-transformed measures; details, including statistics comparing the short and long words, are provided in the appendix.

Procedure

The procedure was the same as in Experiments 3 and 4.

Results and Discussion

The proportion of short and long words correctly placed in order (left panel) and the mean output time (right panel) are shown in Figure 5. As can be seen, there was no word length effect, with recall of short and long words and output times equivalent. These observations were
supported by the following analyses.

The data were analyzed using a two word length (short vs. long) × seven serial positions repeated measures analysis of variance. The main effect of word length was not significant, \( F(1,24) < 1 \). The mean proportion of short words correctly recalled in order was 0.568 (\( SD = 0.167 \)) compared to 0.569 (\( SD = 0.187 \)) for the long words, \( d = 0.006 \). There was a significant main effect of position, \( F(6,144) = 24.857, \ MSE = 0.028, p < 0.01 \). The interaction was not significant, \( F(6,144) = 1.048, \ MSE = 0.014, p > 0.25 \).

The mean output time data were also analyzed using a two word length (short vs. long) × seven serial positions repeated measures analysis of variance. The main effect of length was not significant, \( F(1,24) < 1 \). The mean output time for short words was 1669.4 ms (\( SD = 632.7 \)) compared to 1627.9 ms (\( SD = 668.5 \)) for long words, \( d = 0.064 \). There was a significant main effect of position, \( F(6,144) = 37.233, \ MSE = 1217240.93, p < 0.01 \). The interaction was not significant, \( F(6,144) < 1 \).

Experiment 5 replicated Experiment 4 even though the difference between the short and long words was increased from 1 syllable to 2 syllables: No word length effect was observed, a result predicted by the lexical factors hypothesis but contrary to predictions of the standard model. However, even though the short and long words were equated for various \( z \)-transformed lexical measures as well as on Coltheart’s \( N \), the sets of words did differ significantly by the OLD20 measure and also differed in OLDF. The purpose of Experiment 6 was to assess whether the word length effect remains absent when the short and long words are equated for OLD20 and OLDF.

**Experiment 6**

Experiment 6 compared recall of a 2 and 4 syllable words that were equated on 18
dimensions, including OLD20, the more sensitive measure of orthographic neighborhood size.

The prediction of the standard model is that a word length effect should obtain whereas the prediction from the lexical factors hypothesis is that performance should be equivalent.

Subjects

Twenty-five different volunteers from ProlificAC, with the same inclusion criteria as Experiments 4-6, were paid £10.00 per hour. The mean age was 31.64 ($SD = 5.11$, range 21-38) and 8 identified themselves as female, and 15 identified themselves as male; 2 did not respond to the question.

Stimuli

The stimuli were 36 two-syllable words and 36 four-syllable words. The words were drawn originally from a much larger pool sampled from M. Coltheart (1981), and the original pool was reduced in size until the large and small neighborhood words were equated for concreteness, imageability, and familiarity (from Coltheart); CELEX frequency, orthographic neighborhood size, and frequency of orthographic neighbors (from Medler & Binder, 2005); SUBTLEX$_{us}$ frequency and contextual diversity (from Brysbaert & New, 2009); and OLD20 and OLDF (from Yarkoni et al., 2007). In addition, the short and long words were equated on various $z$-transformed measures, and output time was measured to ensure that was equated. Details, including statistics comparing the short and long words, are provided in the appendix.

Procedure

The procedure was the same as in Experiments 3-5.

Results and Discussion

The proportion of short and long words correctly placed in order (left panel) and the mean output time (right panel) are shown in Figure 6. As can be seen, there was no word length
The data were analyzed using a two word length (short vs. long) × seven serial positions repeated measures analysis of variance. The main effect of word length was not significant, $F(1,24) < 1$. The mean proportion of short words correctly recalled in order was 0.503 ($SD = 0.147$) compared to 0.501 ($SD = 0.169$) for the long words, $d = 0.011$. There was a significant main effect of position, $F(6,144) = 39.042$, $MSE = 0.032$, $p < 0.01$. The interaction was not significant, $F(6,144) = 1.507$, $MSE = 0.018$, $p > 0.15$.

The mean output time data were also analyzed using a two word length (short vs. long) × seven serial positions repeated measures analysis of variance. The main effect of length was not significant, $F(1,24) = 1.257$, $MSE = 160861.69$, $p > 0.25$. The mean output time for short words was 1550.3 ms ($SD = 546.5$) compared to 1598.4 ms ($SD = 586.8$) for long words, $d = 0.085$. There was a significant main effect of position, $F(6,144) = 48.967$, $MSE = 596248.89$, $p < 0.01$. The interaction was not significant, $F(6,144) = 1.014$, $MSE = 207851.57$, $p > 0.40$.

The standard model requires that short words be better recalled than otherwise equivalent long words because more short words can be rehearsed in a given amount of time than long words. Because of the additional rehearsal, short words will decay less than long words and will therefore be better recalled. In contrast, the lexical factors hypothesis predicts a word length effect only when word length is confounded with lexical properties that ordinarily vary with length.

Experiment 7

Experiments 4, 5, and 6 all resulted in finding no difference in recall of short and long words when the short and long words were equated on increasingly more dimensions. It could be
argued that no factor will influence recall performance with so many stringent controls. In other words, our failure to find a word length effect may not be specific to word length, but could be a by-product of the joint influence of all controls. If two sets of words are equated on 18 dimensions, a difference on the 19th dimension may not be sufficient to produce an effect. In order to test this possibility, we selected neighborhood size, one of our lexical factors, and tested its influence when controlling for all other factors, including word length. Specifically, we compared recall of 2-syllable nouns that differed only in orthographic neighborhood size (as measured by Coltheart’s $N$, the $z$-transformed version of this measure, and OLD20). Although we tried to equate the words on the remaining 18 dimensions, and we were able to create a larger pool than in previous experiments, we were unable to equate them for $z$-transformed constrained trigram count.

Subjects

Twenty-five different volunteers from ProlificAC, with the same inclusion criteria as in Experiments 4-7, were paid £10.00 per hour. The mean age was 29.71 ($SD = 5.15$, range 19-38) and 15 identified themselves as female, 8 identified themselves as male, and 1 did not respond to the question.

Stimuli

The stimuli were 70 two-syllable words with from large neighborhoods (as measured by both Coltheart’s $N$ and OLD20), and 70 two-syllable words from small neighborhoods. The words were drawn originally from a much larger pool sampled from M. Coltheart (1981), and the original pool was reduced in size until the large and small neighborhood words were equated for concreteness, imageability, familiarity, number of letters, number of phonemes, and number of syllables (from Coltheart); CELEX frequency and frequency of orthographic neighbors (from
Medler & Binder, 2005); SUBTLEXUS frequency and contextual diversity (from Brysbaert & New, 2009); various z-transformed measures (except the z-transformed constrained trigram count); and OLDF (from Yarkoni et al., 2007). Details, including statistics comparing the two sets of words, are provided in the appendix.

Procedure

The procedure was the same as in Experiments 3-6.

Results and Discussion

The proportion of small and large neighborhood words correctly placed in order (left panel) and the mean output time (right panel) are shown in Figure 7. As can be seen, words from a large neighborhood were better recalled than words from a small neighborhood, even though output times were equivalent. These observations were supported by the following analyses.

The data were analyzed using a two neighborhood size (small vs. large) × seven serial positions repeated measures analysis of variance. The main effect of neighborhood size was significant, \( F(1,24) = 6.609, MSE = 0.037, p < .05 \). The mean proportion of large neighborhood words correctly recalled in order was 0.649 (SD = 0.145) compared to 0.597 (SD = 0.186) for small neighborhood words, \( d = 0.316 \). There was a significant main effect of position, \( F(6,144) = 32.246, MSE = 0.028, p < 0.01 \). The interaction was not significant, \( F(6,144) < 1 \).

The mean output time data were also analyzed using a two neighborhood size (small vs. large) × seven serial positions repeated measures analysis of variance. The main effect of neighborhood size was not significant, \( F(1,24) < 1 \). The mean output time for large neighborhood words was 1543.3 ms (SD = 700.6) compared to 1508.6 ms (SD = 637.0) for small neighborhood words, \( d = 0.052 \). There was a significant main effect of position, \( F(6,144) = 18.871, MSE = 1657394.60, p < 0.01 \). The interaction was not significant, \( F(6,144) < 1 \).
The neighborhood size effect seen in Experiment 7 was quite small in terms of absolute proportion correct (0.649 vs. 0.597), as was the effect size, $d = 0.316$. The primary difference between these results and the ones reported previously are that the stimuli in the current experiment are equated on far more dimensions, including output time. It therefore should not be surprising that once stimuli are equated on so many dimensions, the effect of any one dimension becomes smaller.

General Discussion

Although there exist numerous studies showing that on immediate tests short words are recalled better than long words, there have been a number of reports suggesting that the results hinge on the particular characteristics of the stimuli used, as detailed in the introduction (see Bireta et al., 2006; Neath et al., 2003). In order to examine stimulus properties, we wanted to use large stimulus sets to better minimize the contribution of any one unusual or odd word. However, a number of researchers have argued that the word length effect seen with large stimulus sets differs from the word length effect seen with small stimulus sets. Experiments 1 and 2 demonstrated that the preserved word length effect seen with large stimulus sets under concurrent articulation is eliminated once the short and long words are more fully equated. Moreover, they show that even with a preserved word length effect, word length interacts with concurrent articulation just as with small stimulus sets. This demonstration obviates the need for the ad hoc explanations previously offered for word length effects seen with large stimulus sets, and allows the standard model of decay offset by rehearsal to account for word length effect findings regardless of set size. More importantly, these results also allow us to use large set sizes and have the results apply to prior results from both small and large sets.

Jalbert et al. (2011a) noted that all published studies of the word length effect that
detailed the stimuli used confounded length with neighborhood size, with short words having more neighbors and long words having fewer neighbors. Whereas Jalbert et al. focused on orthographic neighborhood size, short and long words can vary systematically on other lexical dimensions. The lexical factors hypothesis states that past demonstrations of word length effects were really demonstrations of uncontrolled lexical factors that vary with length. A strong prediction of this hypothesis is that when a set of short and long words are equated on all relevant dimensions other than length, there will be no word length effect. Although Jalbert et al. gave two such demonstrations, their stimuli were not fully equated. First, the short and long words differed on the $z$-transformed measure of orthographic neighborhood size (Storkel, 2004). Second, they also differed on OLD20 and OLDF (Yarkoni, et al. 2008).

Experiments 4, 5, and 6 demonstrated that the word length effect observed in Experiment 3, when length was confounded with a number of lexical variables, disappeared once the short and long words were more fully equated. In Experiments 4 and 5, the short and long words differed only in OLD20 and OLDF (in addition to length), and in Experiment 6, the short and long words were equated on 18 dimensions. No word length effects were observed. One possibility is that when words are equated on so many dimensions, the single dimension on which the stimuli do vary is no longer sufficient to affect performance. Experiment 7 tested this idea and found that a neighborhood size effect – albeit a small one – is observable under such conditions.

There are a number of caveats to our results. One caveat is that it is likely that additional confounds are present. For example, in Experiment 7, the small and large neighborhood words were not equated on the $z$-transformed constrained trigram count measure, and there are surely other dimensions that we did not take into account. It could be the case that
on one of these as yet-to-be-identified dimensions, the long words in Experiments 3-6 enjoy an advantage over the short words, and it is therefore this unknown dimension that is driving the results. While we cannot rule out this possibility, we think there is substantial evidence in favor of the lexical factors hypothesis. For example, there are now five demonstrations of a failure to find a word length effect when short words are equated on a number of lexical dimensions (Jalbert et al., 2011, Experiments 3 and 4; and Experiments 4, 5, and 6 of this paper). Together, these studies question the standard model and any theory based on the central idea of decay offset by rehearsal.

A second caveat is that some theorists argue that these results do not apply to their account because open sets were used. For example, Hughes, Marsh, and Jones (2009) distinguish between what they term pure serial recall, in which the same set of items from a closed pool is used on every trial, and nonpure serial recall, in which new items are used on every trial. In pure serial recall, the “burden of processing lies with reproducing the order of the items” (p. 1412) because item information is controlled. It is only when nonpure serial recall is used that lexical and other long-term memory factors affect performance. Baddeley (2012) makes a similar point. One problem with this pure serial recall account is that there are a number of empirical findings that directly challenge these statements. Using closed sets, Walker and Hulme (1999) found that concrete words were better recalled than abstract words, Roodenrys and Quinlan (2000) found that high frequency words were recalled better than low frequency words, and Monnier and Syssau (2008) found that pleasant words were recalled better than neutral words. Put another way, concreteness, frequency, and pleasantness are all long-term factors and all are observable in what Hughes et al. term “pure serial recall”. Second, we note that not all

---

6 We thank a reviewer for raising this point.
versions of the standard model agree with the pure serial recall account. For example, Cowan’s (1999) embedded processes model assumes that working memory is the activated part of long-term memory and therefore long-term effects can effect working memory regardless of the set size.

A third caveat is that while we include orthographic factors in our lexical factors account, we have ignored phonological factors. This is not to say that phonological factors are not important, but rather, we focus on orthographic rather than phonological factors for simplicity. For the former, all one needs to know is that the subject is native speaker of English. For the latter, each dialect and region may have different pronunciation and therefore may differ subtly in phonological neighborhoods. We predict that if the same experiments were re-run using phonological rather than orthographic neighborhoods, the same results would be obtain (assuming the phonological neighborhoods were computed specifically for the particular dialect being tested). Evidence supporting this prediction includes the results of Roodenrys et al. (2002): they actually did investigate phonological rather than orthographic neighborhood size, but because orthographic and phonological neighborhoods are so highly interrelated, their stimuli varied systematically in orthographic neighborhood size as well as phonological neighborhood. More recently, Clarkson, Roodenrys, Miller, and Hulme (2017) replicated the results of Jalbert et al. (2011b), who compared recall of lists that contained both small and large neighborhood words. Whereas Jalbert et al. used measures of orthographic neighborhood size, Clarkson et al. used measures of phonological neighborhood size. Thus, while we have focused on orthographic factors, phonological factors are likely important contributors as well.

In principle it is easy to test which account, the standard model or the lexical factors hypothesis, best predicts when one will observe a word length effect. Because the lexical factors
hypothesis predicts no word length effect when the short and long words are controlled for variations on dimensions other than length, all a researcher need do is demonstrate that word length effects exist when all other potential lexical factors are controlled for (i.e., equating the short and long words for the 18 dimensions we have focused on, and preferably also including phonological dimensions as well). This is certainly a possibility, because as noted above, we have undoubtedly overlooked one (or more) factors, and these might be what is preventing a true effect of length to emerge.

If length does not cause the word length effect, how might a lexical factor affect memory performance? Although we note that other factors are likely involved, we sketch out two possibilities using orthographic neighborhood size and orthographic frequency. We focus on these factors for three reasons. First, there are a growing number of studies that replicate these effects using a variety of different stimuli (e.g., Allen & Hulme, 2006; Derraugh et al., 2017; Glanc & Greene, 2012; Jalbert et al., 2011a; Roodenrys et al., 2002), and second, there is also evidence that both effects are eliminated by concurrent articulation (Jalbert et al., 2011b; Neath et al., 2016). Finally, both lend themselves to a similar theoretical explanation, although this does not rule out other factors or other potential combinations of factors.

There are at least two accounts of word length effects that are not based on the standard model and, in particular, have no role for decay. The first is based on the Feature Model and invokes an interference-based explanation in which short and long words are assumed to differ in the number of segments (Neath & Nairne, 1995). The idea is that, like words in a list, segments of words need to be recalled in the correct order. It is assumed that the probability of a segment assembly error is the same for all segments; given that long words had more segments than short words, short words are more likely to be correctly assembled than long words. The word length
effect is reduced or eliminated by concurrent articulation through interference. Concurrent articulation is seen as adding noise to the representation (see also Murray, Rowan, & Smith, 1988), which affects short words more than long words. The reason is that if noise is added to a word that has not been assembled correctly, it has essentially no effect, whereas if noise is added to a word that has been assembled correctly, it has a much larger effect. This specific account is no longer viable if length \textit{per se} does not cause the word length effect. However, because the segment assembly process was used only for word length effects, removing this does not compromise any other part of the model.

The Feature Model could be revised, in principle, to explain the beneficial effects of neighborhood size and frequency. For both, one could incorporate Roodenrys’ (2009) idea, discussed above, of feedback from neighbors arising from an interactive network, with more frequent neighbors providing more feedback. In the Feature Model, the feedback would selectively offset the loss of individual features that occurs to feature overwriting, which would increase recall. Concurrent articulation would reduce or eliminate these effects in the same way it reduces or eliminates the acoustic confusion effect: Concurrent articulation adds noise, which would offset the advantage of restored features.

The second account of the word length effect that does not include decay focuses on relative distinctiveness. According to SIMPLE (Neath & Brown, 2006, p. 223; see also Hulme et al., 2004; Hulme et al., 2006), “short words are assumed to be more distinctive (i.e., easier to apprehend) due to the less complex phonological information.” Again, this assumption may no longer be warranted if word length \textit{per se} is not a causal factor. Removing this assumption affects only explanations of the word length effect and is therefore not a critical issue for the model. Like the case with the Feature Model, SIMPLE could, in principle, include the feedback
idea. For example, the feedback could selectively enhance the distinctiveness of the large neighborhood words relative to the small neighborhood words, thus rendering them relatively more distinct. More frequent neighbors could also provide more feedback than less frequent neighbors. Concurrent articulation has not yet been incorporated into SIMPLE, but Neath and Brown (2006) speculated about a process similar to that used in the Feature Model in which concurrent articulation is seen as adding noise and therefore reducing relative distinctiveness.

Conclusions

We have found evidence that concurrent articulation interacts with word length in the same way regardless of whether a small or large stimulus set size is used, which means that a single theoretical account can apply to word length effects observed with large or small set sizes. Although we have not resolved the question of what causes the word length effect, we have found additional evidence against the standard model and more evidence in favor of the lexical factors hypothesis. In general, we find that the more lexical factors that are equated for short and long words, the less likely a word length effect will be observed. Therefore, the lexical factors hypothesis is currently a better predictor of when word length effects will be observed.
Author Notes

The research was supported, in part, by grants from the Natural Sciences and Engineering Research Council of Canada to IN, JS-A, and AMS. Portions of this research (Experiments 1 and 2) were presented at the 26th Annual meeting of the Canadian Society for Brain, Behaviour, and Cognitive Science, Ottawa, June 2016. Correspondence may be addressed to ineath@mun.ca.
References


doi:10.3758/BF03193866

doi:10.3758/BRM.41.4.977

doi:10.1037/0033-295X.106.3.551

doi:10.1037/h0056639


doi:10.1080/09658211.2016.1179330


Appendix

Descriptive information about the stimuli used in each experiment. For all tables: CNC = concreteness; FAM = familiarity; IMG = imageability; NLET = number of letters; NPHN = number of phonemes; NSYL = number of syllables (from M. Coltheart, 1981). FREQ = CELEX frequency; Orth = number of orthographic neighbors (Coltheart’s N); OrthF = frequency of the orthographic neighbors (from Medler & Binder, 2005); OrthZ = a z score based on ORTH; C1Z, C2Z, C3Z = z scores based on constrained unigram, bigram, and trigram counts; U1Z, U2Z, U3Z = z scores based on unconstrained unigram, bigram, and trigram counts (computed by the authors); LgWF = log base 10 frequency in SUBTLEXUS corpus; LgCD = log base 10 contextual diversity in SUBTLEXUS corpus (from Brysbaert & New, 2009); OLD20 = orthographic Levenshtein distance; OLDF = frequency of the Levenshtein neighbors (from Yarkoni et al., 2008). N indicates the number of words for which that measure was available, and t and p are the t-test value and probability of whether the two sets of words differ on that measure.
Table A1: The 81 short (1 syllable) and 81 long (2-5 syllable) words from Experiment 1. Shaded cells show significant differences.

|       | CNC | FAM | IMG | NLET | NPHN | NSYL | FREQ | Orth | OrthZ | OrthF | C1Z  | C2Z  | C3Z  | U1Z  | U2Z  | U3Z  | LgWF | LgCD | OLD2O | OLD1F |
|-------|-----|-----|-----|------|------|------|------|------|-------|-------|------|------|------|------|------|------|------|-------|-------|
| Short |     |     |     |      |      |      |      |      |       |       |      |      |      |      |      |      |      |       |       |
| Mean  | 558.94 | 502.58 | 569.53 | 4.23 | 3.40 | 1    | 48.70 | 8.37 | 0.10  | 67.53 | -0.03 | 0.04 | 0.21 | 0.10 | 0.04 | -0.04 | 2.88 | 2.62 | 1.50  | 8.14  |
| SD    | 37.76 | 85.23 | 45.65 | 0.81 | 0.75 | 0    | 77.58 | 5.38 | 0.78  | 169.88 | 0.82 | 0.84 | 0.83 | 0.97 | 1.00 | 0.73 | 0.71 | 0.63 | 0.29  | 0.52  |
| Min   | 395  | 224  | 397  | 3    | 1    | 1    | 0.54  | 0.00 | -1.21 | 0.00  | -2.21 | -1.51 | -0.94 | -2.22 | -1.34 | -0.68 | 1.30 | 1.08  | 1.00  | 6.93  |
| Max   | 617  | 645  | 634  | 6    | 5    | 1    | 435.73 | 22.00 | 2.22  | 1447.35 | 1.89 | 2.02 | 4.42 | 2.50 | 3.04 | 4.84 | 4.45 | 3.79 | 2.40  | 9.22  |
| N     | 80   | 80   | 80   | 81   | 81   | 81   | 81    | 81   | 81    | 81    | 81    | 81    | 81    | 81    | 81    | 81    | 81    | 81    | 81    |

| Long  |     |     |     |      |      |      |      |      |       |       |      |      |      |      |      |      |      |       |       |
| Mean  | 564.27 | 498.71 | 567.09 | 8.60 | 7.43 | 3.19 | 20.77 | 0.31 | -0.28 | 3.46  | -0.87 | -0.69 | -0.51 | -0.09 | -0.26 | -0.39 | 2.36 | 2.16  | 3.24  | 6.57  |
| SD    | 36.76 | 66.94 | 55.04 | 1.45 | 1.23 | 0.53 | 31.94 | 0.58 | 0.74  | 10.64 | 0.72 | 0.52 | 0.45 | 0.76 | 0.77 | 0.55 | 0.76 | 0.68 | 0.65  | 0.57  |
| Min   | 460  | 318  | 409  | 5    | 5    | 2    | 0.12  | 0    | -0.84 | 0.00  | -2.32 | -1.42 | -1.33 | -2.25 | -1.86 | -1.18 | 0.48 | 0.48  | 2.15  | 5.08  |
| Max   | 631  | 641  | 631  | 12   | 11   | 5    | 159.80 | 2    | 3.29  | 49.68 | 0.69 | 1.45 | 2.14 | 1.91 | 1.67 | 1.83 | 3.90 | 3.47 | 4.70  | 7.69  |
| N     | 78   | 78   | 78   | 81   | 81   | 81   | 81    | 81   | 81    | 81    | 81    | 81    | 81    | 81    | 81    | 81    | 81    | 81    | 80    | 80    |

Short Words:

- air, bar, bard, beast, blood, bloom, board, boss, bowl, boy, brain, breeze, bronze, brute, camp, cane, cash, cell, chief, child, church,
claw, coast, code, coin, cord, corn, crag, dawn, dell, dirt, dove, dust, fowl, fur, gem, gift, girl, gold, golf, ink, inn, jail, keg, lad, lark, lice, limb, lime, lump, moss, noose, nun, oats, peach, pipe, pole, queen, rod, sauce, sea, seat, sky, slave, slush, soil, spray, square, stain, star, steam, storm, street, string, stub, suds, tank, toast, tool, vest, wife

Long Words:
accordion, acrobat, admiral, alcohol, alligator, ambassador, ambulance, amplifier, animal, appliance, automobile, avenue, bacteria, beverage, butterfly, caterpillar, cobblestone, colony, committee, daffodil, decoration, diamond, elephant, evangelist, factory, furniture, gallery, gentlemen, grandmother, headquarters, hospital, hurricane, infirmary, inhabitant, instructor, instrument, islander, kerosene, lemonade, library, macaroni, magazine, malaria, material, medallion, microscope, mosquito, musician, newspaper, nursery, officer, opium, orchestra, passageway, performer, photograph, physician, pianist, policeman, potato, prisoner, procession, professor, property, proprietor, prosecutor, refrigerator, restaurant, retailer, revolver, strawberry, tablespoon, tobacco, umbrella, university, utensil, vaccination, vegetable, vehicle, volcano, wholesaler
Table A2: The 101 short (1 syllable) and 101 long (2-4 syllable) words from Experiment 2. Shaded cells show significant differences.

<table>
<thead>
<tr>
<th></th>
<th>CNC</th>
<th>FAM</th>
<th>IMG</th>
<th>NLET</th>
<th>NPHN</th>
<th>NSYL</th>
<th>FREQ</th>
<th>Orth</th>
<th>OrthZ</th>
<th>OrthF</th>
<th>C1Z</th>
<th>C2Z</th>
<th>C3Z</th>
<th>U1Z</th>
<th>U2Z</th>
<th>U3Z</th>
<th>LgWF</th>
<th>LgCD</th>
<th>OLD20</th>
<th>OLDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>460.65</td>
<td>399.79</td>
<td>439.29</td>
<td>4.45</td>
<td>3.43</td>
<td>1</td>
<td>2.54</td>
<td>6.74</td>
<td>-0.08</td>
<td>78.78</td>
<td>-0.25</td>
<td>-0.08</td>
<td>0.01</td>
<td>0.08</td>
<td>-0.05</td>
<td>-0.05</td>
<td>1.60</td>
<td>1.45</td>
<td>1.62</td>
<td>8.09</td>
</tr>
<tr>
<td>SD</td>
<td>81.87</td>
<td>83.68</td>
<td>82.64</td>
<td>0.81</td>
<td>0.75</td>
<td>0</td>
<td>2.20</td>
<td>5.76</td>
<td>0.87</td>
<td>259.00</td>
<td>0.84</td>
<td>0.89</td>
<td>0.81</td>
<td>0.95</td>
<td>0.91</td>
<td>0.66</td>
<td>0.29</td>
<td>0.30</td>
<td>0.33</td>
<td>0.52</td>
</tr>
<tr>
<td>Min</td>
<td>267</td>
<td>204</td>
<td>218</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0.00</td>
<td>0.00</td>
<td>-1.67</td>
<td>0.00</td>
<td>-1.90</td>
<td>-2.12</td>
<td>-1.46</td>
<td>-2.52</td>
<td>-1.57</td>
<td>-0.75</td>
<td>1.00</td>
<td>0.70</td>
<td>1.00</td>
<td>6.71</td>
</tr>
<tr>
<td>Max</td>
<td>606</td>
<td>567</td>
<td>573</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>11.30</td>
<td>22.00</td>
<td>2.22</td>
<td>2184.92</td>
<td>1.72</td>
<td>2.02</td>
<td>2.16</td>
<td>2.41</td>
<td>2.70</td>
<td>3.51</td>
<td>2.00</td>
<td>1.89</td>
<td>2.65</td>
<td>9.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>1.48</td>
<td>0.83</td>
<td>0.09</td>
<td>23.28</td>
<td>28.16</td>
<td>83.14</td>
<td>1.05</td>
<td>11.25</td>
<td>2.16</td>
<td>3.01</td>
<td>5.19</td>
<td>4.91</td>
<td>4.35</td>
<td>1.14</td>
<td>1.88</td>
<td>2.05</td>
<td>1.30</td>
<td>0.44</td>
<td>20.88</td>
<td>18.36</td>
</tr>
<tr>
<td>p</td>
<td>0.14</td>
<td>0.41</td>
<td>0.93</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.30</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Long</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>439.28</td>
<td>389.64</td>
<td>440.50</td>
<td>8.52</td>
<td>7.49</td>
<td>3.02</td>
<td>3.00</td>
<td>0.26</td>
<td>-0.33</td>
<td>1.31</td>
<td>-0.82</td>
<td>-0.62</td>
<td>-0.42</td>
<td>-0.07</td>
<td>-0.28</td>
<td>-0.26</td>
<td>1.55</td>
<td>1.43</td>
<td>3.29</td>
<td>6.58</td>
</tr>
<tr>
<td>SD</td>
<td>119.53</td>
<td>90.07</td>
<td>108.61</td>
<td>1.57</td>
<td>1.24</td>
<td>0.24</td>
<td>3.89</td>
<td>0.58</td>
<td>0.77</td>
<td>5.59</td>
<td>0.74</td>
<td>0.65</td>
<td>0.56</td>
<td>0.87</td>
<td>0.79</td>
<td>0.75</td>
<td>0.26</td>
<td>0.25</td>
<td>0.73</td>
<td>0.64</td>
</tr>
<tr>
<td>Min</td>
<td>237</td>
<td>154</td>
<td>164</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>-1.41</td>
<td>0.00</td>
<td>-2.41</td>
<td>-1.48</td>
<td>-1.20</td>
<td>-2.16</td>
<td>-1.38</td>
<td>1.00</td>
<td>0.90</td>
<td>1.80</td>
<td>4.77</td>
<td></td>
</tr>
<tr>
<td>Max</td>
<td>660</td>
<td>574</td>
<td>649</td>
<td>12</td>
<td>10</td>
<td>4</td>
<td>32.84</td>
<td>2</td>
<td>3.29</td>
<td>42.86</td>
<td>1.00</td>
<td>2.40</td>
<td>2.50</td>
<td>1.81</td>
<td>2.38</td>
<td>2.28</td>
<td>2.00</td>
<td>1.93</td>
<td>4.85</td>
<td>8.70</td>
</tr>
</tbody>
</table>

**Short Words:**

bard, beech, berth, bib, boll, brim, broil, caste, chrome, clang, cleat, clove, croak, dirge, dolt, drape, elm, ewe, feat, flue, foal, foil,
fowl, friar, frock, gasp, germ, gist, glare, gloom, gnat, graft, graph, grate, groan, hale, haze, hue, hymn, jolt, keel, knoll, lark, latch, lathe, lien, lute, lymph, mite, mosque, mousse, newt, noun, ode, pail, pall, pane, peer, perch, pew, pore, prong, rite, roe, sash, scorn, seam, seer, shale, shear, shriek, shrug, sine, sleet, slough, slush, soot, sop, speck, sprint, spruce, stair, starch, steppe, strut, suds, suede, swarm, thaw, throng, thud, urn, vane, verb, volt, wad, watt, welt, yawn, yoke, yore

Long words:
accordance, acrobat, aggressor, albatross, alchemist, algebra, alkali, allotment, amity, ammonia, amplifier, anecdote, apathy, aperture, aptitude, assortment, attribute, bereavement, bivouac, camouflage, candlelight, caravan, carnation, cauliflower, centennial, charlatan, chinchilla, citation, clemency, combustion, commencement, competence, composure, conjunction, connoisseur, contraction, cranberry, daffodil, dandelion, debacle, derelict, distortion, emporium, enamel, enigma, epitaph, etiquette, evergreen, exclusion, exhaustion, expulsion, fallacy, firmament, gardenia, gingerbread, grasshopper, heresy, honeycomb, humankind, incursion, inferno, invader, iota, kerosene, labyrinth, lecturer, magnesium, malady, malaria, mastery, medallion, midshipman, molecule, mosquito, multitude, ornament, paradox, peacemaker, polio, pollution, procession, promenade, reckoning, rectangle, refinement, regency, retention, rhapsody, rosary, scavenger, sobriety, somersault, sonata, spatula, suppression, sycamore, theologian, traveller, underbrush, upheaval, vigilance
Table A3: The 36 short (2 syllable) and 36 long (3 syllable) words from Experiment 3. Shaded cells show significant differences.

<table>
<thead>
<tr>
<th></th>
<th>CNC</th>
<th>FAM</th>
<th>IMG</th>
<th>NLET</th>
<th>NPHN</th>
<th>NSYL</th>
<th>FREQ</th>
<th>Orth</th>
<th>OrthZ</th>
<th>OrthF</th>
<th>C1Z</th>
<th>C2Z</th>
<th>C3Z</th>
<th>U1Z</th>
<th>U2Z</th>
<th>U3Z</th>
<th>LgWF</th>
<th>LgCD</th>
<th>OLD20</th>
<th>OLDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>436.61</td>
<td>455.50</td>
<td>465.06</td>
<td>6.53</td>
<td>5.22</td>
<td>2</td>
<td>18.01</td>
<td>1.75</td>
<td>-0.07</td>
<td>5.50</td>
<td>-0.60</td>
<td>-0.58</td>
<td>-0.41</td>
<td>0.28</td>
<td>-0.07</td>
<td>-0.09</td>
<td>2.30</td>
<td>2.13</td>
<td>2.18</td>
<td>7.10</td>
</tr>
<tr>
<td>SD</td>
<td>103.67</td>
<td>84.61</td>
<td>75.50</td>
<td>0.51</td>
<td>0.90</td>
<td>0.00</td>
<td>22.29</td>
<td>1.52</td>
<td>0.62</td>
<td>8.37</td>
<td>0.45</td>
<td>0.36</td>
<td>0.20</td>
<td>0.80</td>
<td>0.59</td>
<td>0.48</td>
<td>0.66</td>
<td>0.64</td>
<td>0.35</td>
<td>0.51</td>
</tr>
<tr>
<td>Min</td>
<td>257</td>
<td>252</td>
<td>258</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>0.59</td>
<td>0.00</td>
<td>-0.71</td>
<td>0.00</td>
<td>-1.80</td>
<td>-1.13</td>
<td>-0.80</td>
<td>-1.13</td>
<td>-0.77</td>
<td>0.70</td>
<td>0.70</td>
<td>1.50</td>
<td>5.44</td>
<td></td>
</tr>
<tr>
<td>Max</td>
<td>586</td>
<td>559</td>
<td>579</td>
<td>7</td>
<td>7</td>
<td>2</td>
<td>69.84</td>
<td>7.00</td>
<td>2.49</td>
<td>37.60</td>
<td>0.40</td>
<td>0.47</td>
<td>0.25</td>
<td>2.10</td>
<td>1.76</td>
<td>1.39</td>
<td>3.67</td>
<td>3.38</td>
<td>2.90</td>
<td>8.06</td>
</tr>
<tr>
<td>N</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td></td>
<td>t</td>
<td>0.30</td>
<td>0.02</td>
<td>0.34</td>
<td>7.47</td>
<td>7.66</td>
<td>0.03</td>
<td>6.91</td>
<td>5.52</td>
<td>3.94</td>
<td>1.30</td>
<td>1.18</td>
<td>2.04</td>
<td>1.59</td>
<td>1.28</td>
<td>2.08</td>
<td>0.45</td>
<td>0.40</td>
<td>8.19</td>
<td>1.91</td>
</tr>
<tr>
<td></td>
<td>p</td>
<td>0.77</td>
<td>0.98</td>
<td>0.73</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.98</td>
<td>0.00</td>
<td>0.00</td>
<td>0.20</td>
<td>0.24</td>
<td>0.05</td>
<td>0.12</td>
<td>0.20</td>
<td>0.04</td>
<td>0.65</td>
<td>0.69</td>
<td>0.00</td>
<td>0.06</td>
</tr>
<tr>
<td>Long</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>443.81</td>
<td>455.94</td>
<td>458.78</td>
<td>7.75</td>
<td>6.81</td>
<td>3</td>
<td>17.85</td>
<td>0.00</td>
<td>-0.65</td>
<td>0.00</td>
<td>-0.79</td>
<td>-0.68</td>
<td>-0.51</td>
<td>0.01</td>
<td>-0.24</td>
<td>-0.34</td>
<td>2.22</td>
<td>2.07</td>
<td>2.88</td>
<td>6.89</td>
</tr>
<tr>
<td>SD</td>
<td>99.87</td>
<td>94.01</td>
<td>80.30</td>
<td>0.84</td>
<td>0.86</td>
<td>0.00</td>
<td>21.90</td>
<td>0.00</td>
<td>0.07</td>
<td>0.00</td>
<td>0.70</td>
<td>0.38</td>
<td>0.22</td>
<td>0.68</td>
<td>0.55</td>
<td>0.54</td>
<td>0.68</td>
<td>0.63</td>
<td>0.36</td>
<td>0.42</td>
</tr>
<tr>
<td>Min</td>
<td>269</td>
<td>219</td>
<td>305</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>0.7139</td>
<td>0</td>
<td>-0.71</td>
<td>0.00</td>
<td>-1.89</td>
<td>-1.20</td>
<td>-0.77</td>
<td>-1.45</td>
<td>-1.20</td>
<td>-1.02</td>
<td>0.70</td>
<td>0.70</td>
<td>2.35</td>
<td>6.08</td>
</tr>
<tr>
<td>Max</td>
<td>593</td>
<td>585</td>
<td>588</td>
<td>9</td>
<td>8</td>
<td>3</td>
<td>83.35</td>
<td>0</td>
<td>-0.54</td>
<td>0.00</td>
<td>0.65</td>
<td>0.46</td>
<td>0.33</td>
<td>1.52</td>
<td>0.76</td>
<td>1.62</td>
<td>3.29</td>
<td>3.01</td>
<td>3.90</td>
<td>7.54</td>
</tr>
<tr>
<td>N</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td></td>
</tr>
</tbody>
</table>

**Short Words:**

assent, bandit, blunder, chassis, concept, crisis, defeat, discord, dreamer, dungeon, dweller, expanse, express, failure, fashion,
fielder, goddess, harvest, lesson, madman, margin, message, mortal, narrow, passage, piston, polish, portal, repair, shallow, sultan, traitor, tripod, tumble, whistle

Large Neighborhood Words:
accordion, adjective, amateur, aperture, arrival, benefit, cabinet, caravan, cartilage, chemistry, customer, deposit, disquiet, domicile, dowager, emperor, emulsion, epitaph, harmony, honeymoon, inferno, insurance, library, magazine, mechanic, mercury, musician, orderly, poverty, prejudice, recital, regency, religion, residue, salvation, tradition
Table A4: The 36 short (2 syllable) and 36 long (3 syllable) words from Experiment 4. Shaded cells show significant differences.

<table>
<thead>
<tr>
<th></th>
<th>CNC</th>
<th>FAM</th>
<th>IMG</th>
<th>NLET</th>
<th>NPHN</th>
<th>NSYL</th>
<th>FREQ</th>
<th>Orth</th>
<th>OrthZ</th>
<th>OrthF</th>
<th>C1Z</th>
<th>C2Z</th>
<th>C3Z</th>
<th>U1Z</th>
<th>U2Z</th>
<th>U3Z</th>
<th>LgWF</th>
<th>LgCD</th>
<th>OLD20</th>
<th>OLDF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Short</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>359.44</td>
<td>434.11</td>
<td>398.47</td>
<td>5.56</td>
<td>5.67</td>
<td>2</td>
<td>10.39</td>
<td>0.58</td>
<td>-0.52</td>
<td>2.35</td>
<td>-0.80</td>
<td>-0.68</td>
<td>-0.52</td>
<td>0.17</td>
<td>-0.25</td>
<td>-0.34</td>
<td>2.13</td>
<td>1.99</td>
<td>2.46</td>
<td>7.04</td>
</tr>
<tr>
<td>SD</td>
<td>60.44</td>
<td>56.15</td>
<td>57.71</td>
<td>0.69</td>
<td>0.76</td>
<td>0.00</td>
<td>8.63</td>
<td>0.65</td>
<td>0.27</td>
<td>4.89</td>
<td>0.77</td>
<td>0.53</td>
<td>0.19</td>
<td>0.92</td>
<td>0.88</td>
<td>0.42</td>
<td>0.59</td>
<td>0.57</td>
<td>0.34</td>
<td>0.53</td>
</tr>
<tr>
<td>Min</td>
<td>240</td>
<td>308</td>
<td>287</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>0.89</td>
<td>0.00</td>
<td>-0.84</td>
<td>0.00</td>
<td>-2.02</td>
<td>-1.32</td>
<td>-0.85</td>
<td>-1.50</td>
<td>-1.62</td>
<td>-0.87</td>
<td>0.95</td>
<td>0.85</td>
<td>1.80</td>
<td>5.82</td>
</tr>
<tr>
<td>Max</td>
<td>462</td>
<td>500</td>
<td>481</td>
<td>7</td>
<td>7</td>
<td>2</td>
<td>28.97</td>
<td>2.00</td>
<td>0.20</td>
<td>18.68</td>
<td>0.85</td>
<td>0.93</td>
<td>0.13</td>
<td>2.76</td>
<td>2.11</td>
<td>0.87</td>
<td>3.25</td>
<td>3.07</td>
<td>3.35</td>
<td>8.27</td>
</tr>
<tr>
<td>N</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td></td>
</tr>
</tbody>
</table>

|        |     |     |     |      |      |      |      |      |       |       |     |     |     |     |     |     |      |      |       |      |
| **Long** |     |     |     |      |      |      |      |      |       |       |     |     |     |     |     |     |      |      |       |      |
| Mean  | 349.28 | 433.61 | 393.92 | 7.78 | 6.92 | 3.00 | 10.62 | 0.36 | -0.46 | 1.49 | -0.90 | -0.69 | -0.47 | -0.13 | -0.36 | -0.26 | 2.10 | 1.99 | 2.86 | 6.73 |
| SD    | 70.74 | 67.62 | 39.68 | 0.83 | 0.94 | 0.00 | 9.63 | 0.59 | 0.29 | 4.27 | 0.79 | 0.44 | 0.25 | 0.98 | 0.88 | 0.67 | 0.48 | 0.44 | 0.54 | 0.51 |
| Min   | 250 | 154 | 324 | 7 | 5 | 3 | 1.4873 | 0 | -0.71 | 0.00 | -2.41 | -1.26 | -0.71 | -1.72 | -1.52 | -1.00 | 1.04 | 1.00 | 1.80 | 5.55 |
| Max   | 487 | 500 | 478 | 9 | 9 | 3 | 40.93 | 2 | 0.25 | 21.60 | 0.87 | 0.58 | 0.47 | 1.46 | 2.20 | 1.77 | 2.90 | 2.67 | 4.15 | 7.77 |
| N     | 36 | 36 | 36 | 36 | 36 | 36 | 36 | 36 | 36 | 36 | 36 | 36 | 36 | 36 | 36 | 36 | 36 | 36 | 36|

**Short Words:**

assault, ballot, beloved, caucus, deluge, despot, device, errand, essence, excise, forfeit, gallant, genius, hygiene, impulse, incline,
insight, jargon, mankind, menace, misuse, outcome, outpost, physics, prelude, recruit, retreat, revenge, scholar, thinker, torment, treaty, tribute, triumph, upright, zenith

Long Words:

betrayal, bivouac, boundary, bravery, composure, continent, creator, defiance, dignity, dimension, dismissal, dynasty, edition, episode, etiquette, expulsion, exterior, gravity, homicide, ignition, jealousy, loyalty, mastery, multitude, opponent, overlap, ownership, perjury, pioneer, primary, removal, revenue, sanctity, surrender, tyranny, upheaval
Table A5: The 30 short (1 syllable) and 30 long (3 syllable) words from Experiment 5. Shaded cells show significant differences.

<table>
<thead>
<tr>
<th></th>
<th>CNC</th>
<th>FAM</th>
<th>IMG</th>
<th>NLET</th>
<th>NPHN</th>
<th>NSYL</th>
<th>FREQ</th>
<th>Orth</th>
<th>OrthZ</th>
<th>OrthF</th>
<th>C1Z</th>
<th>C2Z</th>
<th>C3Z</th>
<th>U1Z</th>
<th>U2Z</th>
<th>U3Z</th>
<th>LgWF</th>
<th>LgCD</th>
<th>OLD20</th>
<th>OLDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>459.67</td>
<td>473.83</td>
<td>480.23</td>
<td>5.90</td>
<td>4.23</td>
<td>1</td>
<td>17.76</td>
<td>0.70</td>
<td>-0.63</td>
<td>7.89</td>
<td>-0.87</td>
<td>-0.76</td>
<td>-0.51</td>
<td>-0.17</td>
<td>-0.49</td>
<td>-0.31</td>
<td>2.41</td>
<td>2.22</td>
<td>2.06</td>
<td>7.51</td>
</tr>
<tr>
<td>SD</td>
<td>104.84</td>
<td>74.16</td>
<td>94.26</td>
<td>0.40</td>
<td>0.73</td>
<td>0.00</td>
<td>21.32</td>
<td>0.70</td>
<td>0.32</td>
<td>25.98</td>
<td>0.42</td>
<td>0.37</td>
<td>0.34</td>
<td>0.74</td>
<td>0.78</td>
<td>0.41</td>
<td>0.63</td>
<td>0.58</td>
<td>0.22</td>
<td>0.46</td>
</tr>
<tr>
<td>Min</td>
<td>313</td>
<td>296</td>
<td>315</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>0.95</td>
<td>0.00</td>
<td>-1.59</td>
<td>0.00</td>
<td>-1.90</td>
<td>-2.12</td>
<td>-1.46</td>
<td>-1.86</td>
<td>-1.57</td>
<td>-0.68</td>
<td>0.48</td>
<td>0.48</td>
<td>1.80</td>
<td>6.70</td>
</tr>
<tr>
<td>Max</td>
<td>629</td>
<td>591</td>
<td>618</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>68.54</td>
<td>2.00</td>
<td>-0.13</td>
<td>130.59</td>
<td>0.00</td>
<td>-0.18</td>
<td>0.61</td>
<td>1.44</td>
<td>1.22</td>
<td>1.15</td>
<td>3.39</td>
<td>3.19</td>
<td>2.65</td>
<td>9.03</td>
</tr>
<tr>
<td>N</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>

<p>| | | | | | | | | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Long</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>455.53</td>
<td>464.57</td>
<td>476.17</td>
<td>6.17</td>
<td>6.13</td>
<td>3.00</td>
<td>17.34</td>
<td>0.50</td>
<td>-0.65</td>
<td>6.30</td>
<td>-0.78</td>
<td>-0.72</td>
<td>-0.63</td>
<td>-0.31</td>
<td>-0.09</td>
<td>-0.23</td>
<td>2.24</td>
<td>2.05</td>
<td>2.49</td>
<td>7.16</td>
</tr>
<tr>
<td>SD</td>
<td>117.34</td>
<td>96.29</td>
<td>97.08</td>
<td>0.46</td>
<td>0.68</td>
<td>0.00</td>
<td>26.68</td>
<td>0.68</td>
<td>0.23</td>
<td>27.62</td>
<td>0.56</td>
<td>0.29</td>
<td>0.23</td>
<td>0.92</td>
<td>0.76</td>
<td>0.64</td>
<td>0.72</td>
<td>0.68</td>
<td>0.26</td>
<td>0.31</td>
</tr>
<tr>
<td>Min</td>
<td>270</td>
<td>254</td>
<td>240</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>0.12</td>
<td>0</td>
<td>-0.84</td>
<td>0.00</td>
<td>-1.78</td>
<td>-1.16</td>
<td>-0.97</td>
<td>-2.54</td>
<td>-1.69</td>
<td>-0.88</td>
<td>0.60</td>
<td>0.48</td>
<td>1.90</td>
<td>6.61</td>
</tr>
<tr>
<td>Max</td>
<td>662</td>
<td>620</td>
<td>617</td>
<td>8</td>
<td>7</td>
<td>3</td>
<td>119.16</td>
<td>2</td>
<td>-0.13</td>
<td>149.03</td>
<td>0.44</td>
<td>-0.02</td>
<td>-0.07</td>
<td>1.30</td>
<td>1.24</td>
<td>2.28</td>
<td>3.39</td>
<td>3.18</td>
<td>3.00</td>
<td>7.98</td>
</tr>
<tr>
<td>N</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>

**Short Words:**

breeze, bronze, chrome, corpse, frieze, grudge, hymn, knight, length, morgue, mosque, phrase, pledge, praise, prince, proof, scheme,
search, shield, shrimp, sketch, sponge, squeak, squint, strand, thrill, throat, throng, troupe, wealth

Long Words:

alkali, animal, apricot, aurora, avenue, botany, cabinet, citation, colony, comedy, deputy, empire, gaiety, galaxy, injury, iodine, luxury, malady, memory, mutiny, opinion, poetry, potato, remedy, simile, tomato, tyrant, umpire, verity, violin
Table A6: The 36 short (2 syllable) and 36 long (4 syllable) words from Experiment 6. Shaded cells show significant differences.

<table>
<thead>
<tr>
<th></th>
<th>CNC</th>
<th>FAM</th>
<th>IMG</th>
<th>NLET</th>
<th>NPHN</th>
<th>NSYL</th>
<th>FREQ</th>
<th>Orth</th>
<th>OrthZ</th>
<th>OrthF</th>
<th>C1Z</th>
<th>C2Z</th>
<th>C3Z</th>
<th>U1Z</th>
<th>U2Z</th>
<th>U3Z</th>
<th>LgWF</th>
<th>LgCD</th>
<th>OLD20</th>
<th>OLDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>425.56</td>
<td>451.92</td>
<td>457.36</td>
<td>8.00</td>
<td>6.64</td>
<td>2</td>
<td>15.35</td>
<td>0.25</td>
<td>-0.48</td>
<td>1.89</td>
<td>-0.92</td>
<td>-0.75</td>
<td>-0.50</td>
<td>-0.39</td>
<td>-0.51</td>
<td>-0.45</td>
<td>2.23</td>
<td>2.11</td>
<td>3.04</td>
<td>6.48</td>
</tr>
<tr>
<td>SD</td>
<td>97.64</td>
<td>65.13</td>
<td>76.54</td>
<td>0.68</td>
<td>0.90</td>
<td>0.00</td>
<td>16.22</td>
<td>0.44</td>
<td>0.26</td>
<td>6.24</td>
<td>0.67</td>
<td>0.34</td>
<td>0.20</td>
<td>0.79</td>
<td>0.74</td>
<td>0.36</td>
<td>0.62</td>
<td>0.58</td>
<td>0.50</td>
<td>0.56</td>
</tr>
<tr>
<td>Min</td>
<td>260</td>
<td>273</td>
<td>295</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>0.12</td>
<td>0.00</td>
<td>-0.71</td>
<td>0.00</td>
<td>-2.07</td>
<td>-1.23</td>
<td>-0.71</td>
<td>-2.02</td>
<td>-1.77</td>
<td>-1.08</td>
<td>1.04</td>
<td>1.00</td>
<td>2.00</td>
<td>5.13</td>
</tr>
<tr>
<td>Max</td>
<td>562</td>
<td>566</td>
<td>588</td>
<td>9</td>
<td>8</td>
<td>2</td>
<td>49.74</td>
<td>1.00</td>
<td>0.25</td>
<td>32.54</td>
<td>0.44</td>
<td>0.42</td>
<td>1.15</td>
<td>1.75</td>
<td>0.77</td>
<td>3.14</td>
<td>2.99</td>
<td>4.30</td>
<td>7.84</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>416.86</td>
<td>465.53</td>
<td>458.89</td>
<td>8.39</td>
<td>8.22</td>
<td>4.00</td>
<td>14.44</td>
<td>0.22</td>
<td>-0.36</td>
<td>0.90</td>
<td>-0.62</td>
<td>-0.72</td>
<td>-0.48</td>
<td>-0.43</td>
<td>-0.41</td>
<td>-0.45</td>
<td>2.02</td>
<td>1.89</td>
<td>3.13</td>
<td>6.44</td>
</tr>
<tr>
<td>SD</td>
<td>103.92</td>
<td>74.71</td>
<td>77.82</td>
<td>0.73</td>
<td>0.90</td>
<td>0.00</td>
<td>17.41</td>
<td>0.64</td>
<td>0.64</td>
<td>3.19</td>
<td>0.63</td>
<td>0.32</td>
<td>0.18</td>
<td>0.65</td>
<td>0.83</td>
<td>0.38</td>
<td>0.68</td>
<td>0.64</td>
<td>0.41</td>
<td>0.46</td>
</tr>
<tr>
<td>Min</td>
<td>204</td>
<td>272</td>
<td>326</td>
<td>7</td>
<td>6</td>
<td>4</td>
<td>0.3</td>
<td>0</td>
<td>-0.71</td>
<td>0.00</td>
<td>-2.00</td>
<td>-1.28</td>
<td>-0.79</td>
<td>-1.81</td>
<td>-1.72</td>
<td>-1.14</td>
<td>0.48</td>
<td>0.48</td>
<td>2.40</td>
<td>5.16</td>
</tr>
<tr>
<td>Max</td>
<td>624</td>
<td>591</td>
<td>627</td>
<td>9</td>
<td>10</td>
<td>4</td>
<td>74.37</td>
<td>3</td>
<td>2.44</td>
<td>14.58</td>
<td>0.32</td>
<td>-0.01</td>
<td>-0.10</td>
<td>0.59</td>
<td>1.45</td>
<td>0.50</td>
<td>3.44</td>
<td>3.22</td>
<td>3.90</td>
<td>7.42</td>
</tr>
<tr>
<td>N</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td></td>
</tr>
</tbody>
</table>

**Short Words:**

challenge, champion, childhood, chloride, comfort, concrete, conflict, conquest, courtship, discharge, dungeon, exchange, falsehood,
farewell, footstep, frontage, graphite, hardware, henchman, hostage, household, merchant, mischief, moisture, namesake, nightfall, outbreak, outcome, platform, revenge, stranger, substance, torment, tribute, triumph, twilight

**Long Words:**

academy, acidity, adversity, alimony, alligator, amplifier, asparagus, astronomy, aviator, biology, brutality, ceremony, criticism, dandelion, democracy, discovery, educator, emergency, equality, executive, feudalism, geography, heredity, humanity, infinity, machinery, mahogany, majority, monastery, panorama, rigidity, sobriety, territory, testimony, vegetable, velocity
Table A7: The 70 small and 70 large neighborhood words from Experiment 7. Shaded cells show significant differences.

<table>
<thead>
<tr>
<th></th>
<th>CNC</th>
<th>FAM</th>
<th>IMG</th>
<th>NLET</th>
<th>NPHN</th>
<th>NSYL</th>
<th>FREQ</th>
<th>Orth</th>
<th>OrthZ</th>
<th>OrthF</th>
<th>C1Z</th>
<th>C2Z</th>
<th>C3Z</th>
<th>U1Z</th>
<th>U2Z</th>
<th>U3Z</th>
<th>LgWF</th>
<th>LgCD</th>
<th>OLD20</th>
<th>OLDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>493.87</td>
<td>488.84</td>
<td>504.03</td>
<td>5.86</td>
<td>4.63</td>
<td>2</td>
<td>20.50</td>
<td>1.557</td>
<td>-0.389</td>
<td>39.30</td>
<td>0.013</td>
<td>0.015</td>
<td>-0.176</td>
<td>0.263</td>
<td>0.386</td>
<td>0.166</td>
<td>2.551</td>
<td>2.341</td>
<td>1.957</td>
<td>7.266</td>
</tr>
<tr>
<td>SD</td>
<td>99.21</td>
<td>64.32</td>
<td>88.68</td>
<td>1.00</td>
<td>0.92</td>
<td>0</td>
<td>21.33</td>
<td>0.500</td>
<td>0.439</td>
<td>192.09</td>
<td>0.771</td>
<td>0.794</td>
<td>0.752</td>
<td>0.915</td>
<td>0.918</td>
<td>0.878</td>
<td>0.616</td>
<td>0.582</td>
<td>0.238</td>
<td>0.629</td>
</tr>
<tr>
<td>Min</td>
<td>261</td>
<td>306</td>
<td>327</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>0.30</td>
<td>1</td>
<td>-1.666</td>
<td>0</td>
<td>-1.961</td>
<td>-1.624</td>
<td>-1.463</td>
<td>-1.879</td>
<td>-1.281</td>
<td>-0.845</td>
<td>1.080</td>
<td>0.950</td>
<td>1.5</td>
<td>5.856</td>
</tr>
<tr>
<td>Max</td>
<td>620</td>
<td>598</td>
<td>637</td>
<td>8</td>
<td>7</td>
<td>2</td>
<td>82.34</td>
<td>2</td>
<td>0.252</td>
<td>1596.96</td>
<td>1.414</td>
<td>1.851</td>
<td>2.853</td>
<td>2.370</td>
<td>3.119</td>
<td>4.150</td>
<td>3.608</td>
<td>3.227</td>
<td>2.85</td>
<td>9.417</td>
</tr>
<tr>
<td>N</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td></td>
</tr>
</tbody>
</table>

|     |      |      |      |      |      |      |      |      |       |       |      |      |      |      |      |      |      |      |       |       |
| Long |      |      |      |      |      |      |      |      |       |       |      |      |      |      |      |      |      |      |       |       |
| Mean | 497.89 | 496.81 | 514.64 | 5.67 | 4.60 | 2    | 22.97 | 5.386 | 0.756 | 30.58 | 0.051 | 0.134 | 0.166 | 0.151 | 0.274 | 0.191 | 2.506 | 2.300 | 1.694 | 7.248 |
| SD   | 103.48 | 55.53 | 76.32 | 0.96 | 0.81 | 0    | 35.96 | 1.764 | 0.912 | 76.75 | 0.699 | 0.759 | 0.779 | 0.767 | 0.960 | 0.805 | 0.653 | 0.583 | 0.140 | 0.726 |
| Min  | 252   | 351   | 342   | 4     | 3     | 2    | 0.12  | 4   | -0.866 | 0.49  | -1.728 | -1.577 | -1.204 | -1.866 | -1.190 | -0.795 | 0.778 | 0.778 | 1.25  | 5.648 |
| Max  | 645   | 587   | 641   | 7     | 7     | 2    | 186.03 | 11  | 2.725 | 601.78 | 1.433 | 1.936 | 2.843 | 1.908 | 3.238 | 3.704 | 4.050 | 3.630 | 2.25  | 9.028 |
| N    | 70    | 70    | 70    | 70    | 70    | 70   | 70    | 70   | 70    | 70    | 70    | 70    | 70    | 70    | 70    | 70    | 70    | 70    | 70    |

Small Neighborhood Words:

alley, ally, angel, anger, apple, beret, blister, border, boulder, bucket, burner, butcher, casket, ceiling, chicken, circus, combine,
command, coral, culture, distress, dresser, dropper, duty, easel, ego, essay, ether, gavel, goblet, goddess, harem, hermit, ivy, lawyer, lobster, marble, minor, missile, monkey, motor, neuter, opening, owner, panic, pattern, permit, polish, pollen, prayer, quarter, reaper, rebel, relief, resort, rudder, sandal, satin, seller, shower, silver, stable, supper, supply, thinker, toilet, token, turner, unit, whisper

**Large Neighborhood Words:**
basket, battle, berry, blunder, burrow, cable, cafe, cattle, charter, cleaver, closer, concert, content, convent, cunning, curler, dancer, danger, dial, diet, duel, fable, fairy, fellow, ferry, fever, fiddle, fury, gravel, hatchet, hero, jelly, jingle, kettle, lady, leader, level, lily, manner, master, miner, mister, mixer, narrow, navel, noodle, omen, oven, painter, parish, pickle, pimple, platter, puddle, rattle, riddle, rocket, rumble, sable, saga, sender, slipper, story, ticket, toaster, tower, trailer, tumble, wallet, worker
Figure 1: The proportion of short (1 syllable) and long (2-5 syllable) words correctly recalled in order in the quiet (left panel) and concurrent articulation (right panel) conditions in Experiment 1. The short and long words differ in both frequency and orthographic neighborhood size (see Appendix). Error bars show the standard error of the mean.
Figure 2: The proportion of short (1 syllable) and long (2-4 syllable) words correctly recalled in order in the quiet (left panel) and concurrent articulation (right panel) conditions in Experiment 2. The short and long words are equated for frequency but differ in orthographic neighborhood size (see Appendix). Error bars show the standard error of the mean.
Figure 3: The proportion of short (2 syllable) and long (3 syllable) words correctly recalled in order (left panel) and the mean output time (right panel) in Experiment 3. The short and long words also differ in orthographic neighborhood size (see text for details). Error bars show the standard error of the mean.
Figure 4: The proportion of short (2 syllable) and long (3 syllable) words correctly recalled in order (left panel) and the mean output time (right panel) in Experiment 4. The short and long words are equated for orthographic neighborhood size (see text for details).

Error bars show the standard error of the mean.
Figure 5: The proportion of short (1 syllable) and long (3 syllable) words correctly recalled in order (left panel) and the mean output time (right panel) in Experiment 5. The short and long words are equated for orthographic neighborhood size (see text for details). Error bars show the standard error of the mean.
Figure 6: The proportion of short (2 syllable) and long (4 syllable) words correctly recalled in order (left panel) and the mean output time (right panel) in Experiment 6. The short and long words are equated for OLD20 (see text for details). Error bars show the standard error of the mean.
Figure 7: The proportion of large and small neighborhood words recalled in order (left panel) and the mean output time (right panel) in Experiment 7. Error bars show the standard error of the mean.