# How To: Use the MRC Psycholinguistic Database

URL: <https://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm>

## What is it?

The MRC Database is a searchable dictionary containing over 150,000 words, with linguistic and psycholinguistic attributes for the majority of the words. The MRC Psycholinguistic database can be used to generate word stimuli that fulfil a set of criteria that you have determines, different properties and control for other properties. As an example, you might create a set of stimuli that differ on word frequency but control for imageability.

It can be a challenge to use the database without getting lots of words from the beginning of the alphabet as the results tend to be output in alphabetical order. Here is one example of how to choose word stimuli that are nouns and verbs that differ on concreteness

When designing studies using word stimuli it is very easy to fall into a trap of introducing bias with confounds. We have discussed possible confounds when using word stimuli in Week 3 of the EDP course

## How to Search the MRC

1. **Go to** <https://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm>
2. Select the types of information that you would like your report to show. Some of the parameters are more obvious than others (for a list of the ones that are most commonly used see Appendix1). For this exercise select
3. Word
4. Number of letters
5. Word Frequency
6. Concreteness rating
7. Common part of speech
8. Next, we have to limit the range of words to be output, because the database will output a maximum of 5000 words and we don't want words just from the beginning of the alphabet. Filtering on word length does this. Scroll down to the Optional settings below and in the **NLET min and max values** **type 4 & 5** to find only words that are four or five letters in length



1. Finally, we can filter on word type, because in this case we only want nouns and verbs (you might only want nouns depending upon the results of your literature search)





1. Now click the **Go** button
2. Your output will look something like the image to the ***right***
3. This output is a little unwieldy, so I suggest you select all the words and the parameters in the columns to the right, and paste into an Excel spreadsheet where it will look like the image below



1. To convert the text to columns, select column A, then go to the Data menu from the toolbar, and choose Text to Columns. Doing this will allow you to sort by concreteness values and you can then choose stimuli from the list according to the parameters you have set at the start.

## Appendix 1

This appendix shows the primary variables that may be used in developing stimuli for your Lexical Decision Task. The marking criteria suggests that you choose at least two parameters for your study, based on the literature that you have found.

Number of letters

number of letters in each word, useful for controlling for word length effects

Number of phonemes

number of phonemes, phonemes are perceptually distinct units of sound in a specified language that distinguish one word from another, for example p, b, d, and t in the English words pad, pat, bad, and bat, useful for controlling for phonological similarity effects

Number of syllables

a unit of pronunciation having one vowel sound, with or without surrounding consonants, forming the whole or a part of a word; for example, there are two syllables in water and three in inferno, useful for controlling for syllabic effects

Kucera-Francis written freq

The frequency of a word in English language as given in the norms of Kucera and Francis (1967). The maximum frequency in the file is 69971, the minimum is 0. The more commonly used a word is (the more frequently it occurs, the more likely to be recalled, recognised, read). Bear in mind this was originally published in 1967 and there are possibilities that language has evolved and of course new words have entered the lexicon

Thorndike-Lorge written freq

This is the frequency of occurrence as given in the L count of Thorndike and Lorge (1942). If you plan to use this frequency count, you are advised to read details about it in the Thorndike-Lorge book. For example, the frequency value of a singular word which has a regular plural includes the frequency of the plural form, and this is true for other kinds of derivations too. Again a very old account of written frequency, so keep that in mind

Brown verbal frequency

This stands for the frequency of occurrence in verbal language derived from the London-Lund Corpus of English Conversation by Brown (1984). There are 14529 entries for 8985 different strings in the WORD field. The range of entries is 0 - 6833 with a mean of 35 and a standard deviation of 252

Familiarity rating

This stands for 'printed familiarity'. The FAM values were derived from merging three sets of familiarity norms: Pavio (unpublished), Toglia and Battig (1978) and Gilhooly and Logie (1980). The method by which these three sets of norms were merged is described in detail in Appendix 2 of the MRC Psycholinguistic Database User Manual (Coltheart, 1981a). This method may not meet with everyone's approval. FAM values lie in the range 100 to 700 with the maximum entry of 657, a mean of 488 and a standard deviation of 99: note that they are integer values (in the original norms the equivalent range was 1.00 to 7.00)

Concreteness rating

Concreteness evaluates the degree to which the concept denoted by a word refers to a

perceptible entity such as a physical object in the real world, such as a dog, a ball, or an ice cream cone. An abstract noun refers to an idea or concept that does not exist in the real world and cannot be touched, like freedom, sadness, or permission.

In this database it is derived from a merging of the Pavio, Colerado, and Gilhooly-Logie norms: details of merging are given in Appendix 2 of the MRC Psycholinguistic Database User Manual (Coltheart, 1981a). CONC values are integer, in the range 100 to 700 (min: 158; max 670; mean 438; s.d. 120). Using concreteness, you may want to use words that conform to the mean value of 438 with a range either side of that value, to ensure words are similarly rated for concreteness (opposite = abstract, smaller values indicate lower concreteness, and higher abstract ratings)

Imageability rating

Imageability is a psycholinguistic variable that indicates how well a word gives rise to a mental image or sensory experience, examples might include apple (high imageability) v truth (low imageability)

In this database the ratings are derived from merging the three sets of norms referred to above, and having values in the range 100 to 700 (min 129; max 669; mean 450; s.d. 108). High imageability words are typically processed more quickly than low.

Meaningfulness

These are the meaningfulness ratings from the Toglia and Battig (1978), multiplied by 100 to produce a range from 100 to 700 (min 127; max 667; mean 415; s.d. 78).

Age of Acquisition

This is age of acquisition from the norms of Gilhooly and Logie (1980), multiplied by 100 to produce a range from 100 to 700 (min 125; max 697; mean 405; s.d. 120). More useful when working with child populations as it is helpful to know if a word would typically be known by a child.