Behavior Research Methods, Instruments, & Computers
2004, 36 (3), 384-387

Imageability ratings for 3,000 monosyllabic words

MICHAEL J. CORTESE
College of Charleston, Charleston, South Carolina

and

APRIL FUGETT
Morehead State University, Morehead, Kentucky

Imageability ratings made on a 1–7 scale and reaction times for 3,000 monosyllabic words were obtained from 31 participants. Analyses comparing these ratings to 1,153 common words from Toglia and Battig (1978) indicate that these ratings are valid. Reliability was assessed (α = .95). The information obtained in this study adds to that of other normative studies and is useful to researchers interested in manipulating or controlling imageability in word recognition and memory studies. These norms can be downloaded from www.psychonomic.org/archive/.

The present study provides imageability ratings and imaegability reaction times (RTs) for 3,000 monosyllabic words. The primary purpose of these ratings is to provide researchers with a source of information that can be used to control (Cortese, 1998; Cortese & Simpson, 2000), manipulate (Cortese, Simpson, & Woolsey, 1997; Frost et al., 2003; Glanzer & Adams, 1990; Hintzman, Caulton, & Curran, 1994; Strain, Patterson, & Seidenberg, 1995), or analyze (see, e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004) imageability[1] in word processing and memory studies. In fact, in some previous studies in which imageability was manipulated or controlled for, normative information about their stimuli was collected prior to the main study (see, e.g., Cortese, 1998; Cortese & Simpson, 2000, Zevin & Balota, 2000).

The present study stems from the fact that, for a long time, researchers have been interested in the extent to which meaning influences basic word recognition processes (see, e.g., Balota, 1990) and memory (see, e.g., Paivio, 1971, for a recent view of dual coding theory, and Sadoski & Paivio, 2001, for a discussion of imageability in word processing from that perspective). For example, in studies of reading aloud, Strain and colleagues (Strain & Herdman, 1999; Strain et al., 1995, 2002; see also Cortese et al., 1997) varied imageability along with spelling-to-sound consistency and found that imageability was related to naming latencies for inconsistent words (e.g., *comb* vs. *caste*) but not for consistent words (e.g., *cliff* vs. *cleft*). This result has important implications for models of reading aloud (see, e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Plaut, McClelland, Seidenberg, & Patterson, 1996). Specifically, this result supports the *interactivity assumption* (i.e., that different sources of information—e.g., semantic, phonological, and orthographic sources—can influence the computation of a code at any level) present in contemporary models of reading aloud. More recently, by examining the relationship between consistency and imageability in a neuroimaging study, Frost et al. (2003) have found brain areas that reflect the interaction between phonology and semantics.

However, it is important to note that the interaction between consistency and imageability reported by Strain and colleagues has been the focus of a recent debate. Specifically, Monaghan and Ellis (2002) have argued that there is little evidence of the interaction between imageability and consistency once age of acquisition (AoA) is considered. In fact, Monaghan and Ellis claim that it is AoA, and not imageability, that interacts with consistency. The relationship that AoA has with measures of word recognition has also been scrutinized (see, e.g., Zevin & Seidenberg, 2002). Obtaining imageability estimates for a large number of words will allow for more extensive testing of the interactivity assumption, and it may help resolve the current debate about the relationship between imageability and AoA.

It is important to note that the vast majority of studies in which word processing has been examined have employed the use of monosyllabic words. Interestingly, information about the imageability of (the referents of) these words in English is quite limited. For example, until now, imageability ratings conducted by a common sample of people were available for only a little over one third of our word corpus. Reliance on this subset limits the kinds of studies that can be performed as well as the extent to which megastudies (e.g., Balota et al., 2004) can evaluate the influence of this factor on performance. By obtaining these data, our study greatly reduces this limitation. Also, imageability estimates for this large corpus of monosyllabic words can be used in the selection of stimuli for

This work was partially supported by Kentucky NSF EPSCoR Grant EPS0132295. Correspondence concerning this article should be addressed to M. J. Cortese, Department of Psychology, College of Charleston, 66 George St., Charleston, SC 29424 (e-mail: cortesem@cofc.edu).

other types of experiments involving words. These experiments include explicit and implicit memory experiments, lexical decision experiments, and many more.

It is important to note that our corpus consists of 1,153 words that were rated in Toglia and Battig's (1978) study. This will allow the validity of these norms to be established. Our instructions were very similar to those of Toglia and Battig. However, the ratings will be recorded via a computer keyboard instead of the pencil-and-paper method employed by Toglia and Battig. In addition, although speed will not be emphasized, we will obtain latencies for these ratings.

METHOD

Participants

Thirty-one Morehead State University undergraduates enrolled in a psychology course participated in the study. They received extra credit for their participation.

The Word Corpus

The word corpus consists of 3,000 monosyllabic words (see the electronic archive for ratings and latencies). A good deal of information is already known about a large proportion of this corpus. For example, Balota et al. (2004) have obtained reading-aloud and lexical decision data from younger and older adults on the majority of these items. In addition, objective frequency (Kučera & Francis, 1967) and subjective frequency (Balota, Pilotti, & Cortese, 2001) are available for the majority of items.

Procedure

Ratings were collected via microcomputers in a computer lab in two sessions held 1 week apart. Each session lasted between 1.25 and 2.0 h. One thousand five hundred words were rated in each session. Each session was divided into four blocks of 375 words. Opportunities for breaks occurred at the end of each block. Each participant rated the words in a different random order.

On each trial, a word was presented at the center of the screen. A number keypad on the right side of the keyboard was used to enter each rating. The instructions were essentially the same as those employed by Toglia and Battig (1978) and Paivio, Yuille, and Madigan (1968), with two exceptions. First, in the written instructions (see the Appendix), the example of a highly imageable word was changed from the Colorado mascot (used in Toglia & Battig, 1978) to an eagle (the Morehead State University mascot). Second, the instructions were modified for computer use. For example, the participants were informed verbally that stimuli would appear one at a time at the center of the computer screen, and that their ratings should be entered via the number keypad located on the right side of the keyboard. It is important to note that although RTs to the ratings were measured, the participants were not informed of this and were not instructed to respond as quickly as possible. Because of the number of words that were rated, it was in the participants' best interest to maintain a fairly quick pace. Therefore, the primary purpose of recording RTs was to allow us to eliminate ratings that were made prematurely (i.e., responses taking less than 300 msec were eliminated; see below).

ANALYSES, RESULTS, AND DISCUSSION

First, the data were screened. We eliminated latencies and ratings whenever the rating was made in less than 300 msec. Because speed was not emphasized in our instructions, no upper limit for response latencies was established. Through this screening criterion, 3.7% of the data were eliminated.

One concern might be that the data are noisy due to the number of words that were rated by each participant. To address this concern, we compared the ratings for the 1,153 words common to both our database and Toglia and Battig's (1978) database. For this common set of words, the mean rating from our sample was lower than that of Toglia and Battig's sample [4.57 vs. 4.79; $t(1,152) = 9.21$, $p < .01$]. This is interesting given that our instructions were similar to theirs. More important, our standard deviation (*SD*) was larger than Toglia and Battig's (1.52 vs. .96, respectively). This indicates that our sample was using more of the scale than Toglia and Battig's sample. This characteristic may give these ratings more predictive power than Toglia and Battig's ratings because the range is less restricted. Finally, there was a strong correlation between our ratings and Toglia and Battig's ratings [$r(1,151) = .89$, $p < .01$]. This outcome provides evidence of criterion-related validity. Reliability was assessed via alpha, and $\alpha = .95$. Establishing that these estimates are reliable and valid also further establishes the method employed as a reliable technique for obtaining data for a large corpus of words.

Regarding the imageability latencies, there are two things to note. First, the average latency was 1,423.2 msec ($SD = 208.5$). Second, latencies were negatively correlated with imageability ratings [$r(2,998) = -.47$, $p < .01$]. In other words, the more imageable the referent of the word was, the shorter the RT was. Paivio (1968) obtained a latency of about 1 sec and a correlation between imageability ratings and RTs of -70. The differences between the studies could be due to the different items rated and/or the directions given.

It is also important to note that these norms have been used recently for predicting naming and lexical decision latencies (see Balota et al., 2004). These ratings accounted for unique variance in both tasks.

In summary, we have provided imageability norms for 3,000 words so that researchers may have access to this information for studies that manipulate or control imageability. The comparison of 1,153 items common to our study and Toglia and Battig's (1978) study indicate that our ratings and method are reliable and valid. Finally, these ratings have already been shown to have predictive power in naming and lexical decision performance (Balota et al., 2004). This demonstrates the utility of the norms for gleaning insight into word recognition processes, and the potential that these norms have for other cognitive studies.

REFERENCES

Balota, D. A. (1990). The role of meaning in word recognition. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 9-32). Hillsdale, NJ: Erlbaum.

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

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

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, **108**, 204-256.

Cortese, M. J. (1998). Revisiting serial position effects in reading. *Journal of Memory & Language*, **39**, 652-665.

Cortese, M. J., & Simpson, G. B. (2000). Regularity effects in word naming: What are they? *Memory & Cognition*, **28**, 1269-1276.

Cortese, M. J., Simpson, G. B., & Woolsey, S. (1997). Effects of association and imageability on phonological mapping. *Psychonomic Bulletin & Review*, **4**, 226-231.

Frost, S. J., Mencl, W. E., Sandak, R., Moore, D. L., Mason, S. A., Rueckl, J. G., Katz, L., & Pugh, K. R. (2003). Semantic constraints on spelling-to-sound consistency effects. *Abstracts of the Psychonomic Society*, **8**, 71.

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 5-16.

Hintzman, D. L., Caulton, D. A., & Curran, T. (1994). Retrieval constraints and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 275-289.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Monaghan, J., & Ellis, A. W. (2002). What exactly interacts with spelling–sound consistency in word naming? *Journal of Experiment Psychology: Learning, Memory, & Cognition*, **28**, 123-136.

Paivio, A. (1968). A factor-analytic study of word attributes and verbal learning. *Journal of Verbal Learning & Verbal Behavior*, **7**, 41-49.

Paivio, A. (1971). *Imagery and verbal processes.* New York: Holt, Rinehart, & Winston.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery and meaningfulness values for 925 words. *Journal of Experimental Psychology Monograph Supplement*, **76**(3, Pt. 2).

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

Sadoski, M., & Paivio, A. (2001). *Imagery and text: A dual coding theory of reading and writing*. Mahwah, NJ: Erlbaum.

Strain, E., & Herdman, C. M. (1999). Imageability effects in word naming: An individual differences analysis. *Canadian Journal of Experimental Psychology*, **53**, 347-359.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single word naming. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1140-1154.

Strain, E., Patterson, K., & Seidenberg, M. S. (2002). Theories of word naming interact with spelling–sound consistency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 207-214.

Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms.* Hillsdale, NJ.: Erlbaum.

Zevin, J. D., & Balota, D. A. (2000). Priming of lexical and sublexical pathways during naming. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 121-135.

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in reading and other tasks. *Journal of Memory & Language*, **47**, 1-29.

NOTE

1. Actually, Glanzer and Adams (1990) and Hintzman et al. (1994) manipulated concreteness. Concreteness is highly correlated with imageability (Toglia & Battig, 1978).

ARCHIVED MATERIALS

The following materials may be accessed through the Psychonomic Society's Norms, Stimuli, and Data archive, http://www.psychonomic.org/archive/. To access these files, search the archive for this article using the journal (*Behavior Research Methods, Instruments, & Computers*), the first author's name (Cortese), and the publication year (2004).

File: Cortese-BRMIC-2004.zip.

Description: The compressed archive file contains two files:

cortese2004norms.xls, containing the norms developed in the present article, as a 96K tab-delimited text file generated by Excel 2001, and cortese2004norms.pdf, containing the same norms in PDF format generated by Adobe Acrobat. Each row of each file represents one of 3,000 words; each column, one of six measures (including imageability rating and standard deviation) to describe each word. The description of this file is included at the top of the file.

Author's e-mail address: cortesem@cofc.edu.

Author's Web site: http://www.cofc.edu/~cortesem.

APPENDIX
Written Instructions Used for the Imageability Rating Task

Words differ in their capacity to arouse mental images of things or events. Some words arouse a sensory experience, such as a mental picture or sound, very quickly and easily whereas other words may do so only with difficulty (i.e., after a long delay) or not at all.

The purpose of this experiment is to rate a list of 3,064 words as to the ease or difficulty with which they arouse mental images. Any word, that in your estimation, arouses a mental image (i.e., a mental picture or sound, or other sensory experience) very quickly and easily should be given a high imagery rating (at the upper end of the numerical scale). Any word that arouses a mental image with difficulty or not at all should be given a low imagery rating (at the lower end of the numerical scale). For example think of the word "EAGLE." "EAGLE" would probably arouse an image (e.g., of the Morehead State University mascot) relatively easily and would be rated as high. "RELEVANT" would probably do so with difficulty and be rated as low imagery.

Because words tend to make you think of other words as associates, it is important that your ratings not be based on this, and that you judge only the ease with which you get a mental image of an object or event in response to each word.

Your imagery ratings will be made on a 1 to 7 scale. A value of 1 will indicate a low imagery rating, and a value of 7 will indicate a high imagery rating. Values of 2 to 6 will indicate intermediate ratings.

1	2	3	4	5	6	7
Low			Medium			High

When making your ratings, try to be as accurate as possible, but do not spend too much time on any one word.

If you have any questions, ask the experimenter now. Otherwise, press <ENTER> to begin the experiment.

(Manuscript received December 26, 2003;
revision accepted for publication July 18, 2004.)