The Minho Word Pool: Norms for imageability, concreteness, and subjective frequency for 3,800 Portuguese words

Ana Paula Soares[1] · Ana Santos Costa[1] · João Machado[1] · Montserrat Comesaña[1] ·
Helena Mendes Oliveira[1]

Abstract Words are widely used as stimuli in cognitive research. Because of their complexity, using words requires strict control of their objective (lexical and sublexical) and subjective properties. In this work, we present the Minho Word Pool (MWP), a dataset that provides normative values of imageability, concreteness, and subjective frequency for 3,800 (European) Portuguese words—three subjective measures that, in spite of being used extensively in research, have been scarce for Portuguese. Data were collected with 2,357 college students who were native speakers of European Portuguese. The participants rated 100 words drawn randomly from the full set for each of the three subjective indices, using a Web survey procedure (via a URL link). Analyses comparing the MWP ratings with those obtained for the same words from other national and international databases showed that the MWP norms are reliable and valid, thus providing researchers with a useful tool to support research in all neuroscientific areas using verbal stimuli. The MWP norms can be downloaded along with this article or from http://p-pal.di.uminho.pt/about/databases.

Keywords Imageability · Concreteness · Subjective frequency · Subjective norms · Portuguese

✉ Ana Paula Soares
asoares@psi.uminho.pt

[1] Human Cognition Lab, CIPsi, School of Psychology, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

Words have been used extensively in different areas of cognitive psychology (e.g., language, memory, emotion, and learning). Despite the familiarity we have with words in our everyday lives, they are extremely complex stimuli, and a very strict control of their characteristics is required when using them as experimental materials. Besides form (orthography and phonology) and meaning (semantics), words also have several other lexical and sublexical properties that need to be accurately processed to allow for the correct identification of one particular word within the thousands stored in our mental lexicon, which may only slightly differ, such as the case of *word* versus *work*. Although we are not aware of the complexity of this process—from the moment a child learns how to read, recognizing words becomes a virtually automatic and effortless activity—word recognition is considered one of the most complex activities our cognitive system performs. Trying to "crack the code" has been a main focus of research in cognitive science since its early beginnings.

However, even though we recognize words very efficiently—it is estimated that we take less than a quarter of a second to recognize a word—the vast amount of studies conducted in the last decades have revealed the existence of a number of variables that affect the speed and the accuracy with which words are processed, recognized or recalled. Among these variables are not only words' characteristics that depend on the objective analysis of their proprieties at the lexical and sublexical levels, like word length (e.g., Ferrand et al., 2011; New, Ferrand, Pallier, & Brysbaert, 2006), word frequency (e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Brysbaert et al., 2011; Ferrand et al., 2011), the diversity of contexts in which a word appears (e.g., Adelman, Brown, & Quesada, 2006; Parmentier, Comesaña, & Soares, 2016; Perea, Soares, & Comesaña, 2013), or the orthographic similarity with other words in the lexicon (e.g., Coltheart, Davelaar, Jonasson, & Besner, 1977; Ferrand et al., 2011; Yarkoni, Balota, & Yap,

2008), but also words' characteristics that depend on the personal experiences that individuals had with the use of those words in their language (subjective properties).

This kind of word properties include variables such as word imageability (i.e., the ease and speed with which a word evokes a mental image—e.g., Paivio, Yuille, & Madigan, 1968), concreteness (i.e., the degree to which words refer to objects, persons, places, or things that can be experienced by the senses—e.g., Paivio et al., 1968), experiential familiarity (i.e., the degree to which individuals know and use words in their everyday life—e.g., Gernsbacher, 1984), subjective frequency (i.e., the estimation of the number of times a word is encountered by individuals in its written or spoken form—e.g., Balota, Pilotti, & Cortese, 2001), age of acquisition (AoA; i.e., the estimation of the age at which a word was learned—e.g., Carroll & White, 1973), and also words' affective properties such as the degree of (un)pleasantness (valence) and/or the degree of activation (arousal) a word triggers in individuals (e.g., Osgood, Suci, & Tannenbaum, 1957).

Studies conducted so far with different languages, tasks and paradigms, have shown that, additionally to objective measures like word frequency, contextual diversity, word length, and orthographic similarity, subjective measures like imageability and/or concreteness (e.g., Paivio, 1971, 1986; Schwanenflugel, 1991; Strain & Herdman, 1999), experiential familiarity (e.g., Gernsbacher, 1984; Gilhooly & Logie, 1980; Gordon, 1985), subjective frequency (e.g., Balota et al., 2001; Brysbaert & Cortese, 2011; Thompson & Desrochers, 2009), AoA (e.g., Bird, Franklin, & Howard, 2001; Brysbaert & Cortese, 2011; Ferrand et al., 2011; Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012), and words' affective content (e.g., Altarriba & Bauer, 2004; Altarriba, Bauer, & Benvenuto, 1999; Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011; Vigliocco et al., 2013) also play a role in accounting for significant percentages of variance. Overall, the above-mentioned research suggests that words acquired earlier in life, more concrete (see, however, Kousta et al., 2011, for a concreteness reverse effect), more imaginable, more familiar, rated with higher estimations of use in daily life, more pleasant and/or arousing, are recognized, named, categorized, and recalled more quickly and accurately than words that score lower in these subjective measures, particularly for words with a low frequency of occurrence in a language.

Although these subjective variables are not independent from each other, an increasing body of studies suggest that they theoretically and empirically constitute distinct constructs (e.g., Connell & Lynott, 2012; Dellantonio, Mullatti, Pastore, & Job, 2014; Kousta et al., 2011). For example, although imageability is highly correlated with concreteness (.83 in Paivio's work and around that value in subsequent studies—e.g., Altarriba et al., 1999; Connell & Lynott, 2012; Gilhooly & Logie, 1980; Toglia & Battig, 1978), which has supported the interchangeable use of both measures in experimental research (see, e.g.,

Fliessbach, Weis, Klaver, Elger, & Weber, 2006), a growing body of evidence has suggested that imageability and concreteness capture different word properties (e.g., Connell & Lynott, 2012; Dellantonio et al., 2014; Kousta et al., 2011). In their seminal work, Paivio and collaborators (1968) already acknowledged the differences between these two constructs. Indeed, although words denoting objects experienced by the senses (i.e., concrete words), would reevoke a mental image associated with that sensory experience more easily than abstract words, Paivio et al. highlighted that for certain abstract words, particularly those denoting affective states (e.g., anger, joy), the positive correlation between the two constructs is not observed. Like the vast majority of abstract words (e.g., liberty, justice), affective words score lower in concreteness, but unlike them, they score higher in imageability. This "special status" of emotional words has also been confirmed by other studies (e.g., Altarriba et al., 1999; Altarriba & Bauer, 2004; Kousta et al., 2011; Vigliocco et al., 2013), which have supported the assertion that emotional words are represented and processed differently from other word types stored in our lexicon.

In a recent study aiming to disentangle the two dominant accounts of the differences between concrete and abstract words—the dual-coding theory of Paivio (1971, 1986) and the context availability model of Schwanenflugel and colleagues (Schwanenflugel, 1991; Schwanenflugel & Shoben, 1983), Kousta et al. (2011) demonstrated that concreteness and imageability cannot be understood as the same underlying construct. Specifically the analyses conducted on the distribution of the data, revealed that whereas concreteness showed a bimodal distribution—with two different modes capturing the ontological distinction between spatiotemporally bounded (concreteness) and non-spatiotemporally-bounded (abstractness) concepts, imageability ratings showed a unimodal distribution instead, indexing the amount of sensory information associated with the words. In this context, and motivated by the embodied theories of cognition (e.g., Barsalou, Santos, Simmons, & Wilson, 2008; Vigliocco, Meteyard, Andrews, & Kousta, 2009), some authors have suggested alternative measures to differentiate these constructs such as the strength of perceptual experience (Connell & Lynott, 2012), the mode of acquisition (MoA; Della Rosa, Catricalà, Vigliocco, & Cappa, 2010) or the sensory experience rating (SER; Juhasz, Yap, Dicke, Taylor, & Gullick, 2011), variables aiming to assess the degree to which a word evokes sensory/perceptual experiences. As claimed by Barsalou and others (e.g., Barsalou et al., 2008; Dellantonio et al., 2014; Kousta et al., 2011; Vigliocco et al., 2009), although both concrete and abstract concepts are represented as situated simulations (i.e., a partial reenactment of the perceptual, motor, and affective neural activation experienced during the acquisition of those concepts), they differ in the focal content of the situations they applied to. Although concrete concepts are represented through a narrow range of situations relying mainly on the perceptual and motor neural information, abstract concepts are more multimodal

since they rely on the social, introspective and affective neural information represented in a wide range of situations, which can also help to explain the "special status" of emotional words on those ratings.

Indeed, this was the interpretation advanced by Kousta et al. (2011) for the abstractness effect observed (i.e., faster response times for abstract than concrete words), a result that collides with the vast amount of studies showing that concrete words are recognized, named, and recalled more easily than abstract words (e.g., Fliessbach et al., 2006; Schwanenflugel & Shoben, 1983). Note that this abstractness effect is not accounted for by the dual-coding theory of Paivio or by the context availability model of Schwanenflugel and colleagues, since both predict that concrete words will be processed and recalled better than abstract words, although using different underlying mechanisms.[1] Kousta et al. (2011) suggested that since abstract words were more strongly associated with affective states than concrete words (i.e., they were more "affectively valenced"), these denser affective associations could explain why abstract words were more easily recognized than concrete words when all objective and subjective word properties known to affect word recognition (including familiarity, imageability, and context availability) were controlled for. Nevertheless, studies aiming to test directly how emotionality relates to other subjective measures (e.g., imageability familiarity, AoA subjective frequency) are still lacking in literature.

Subjective frequency is also strongly related with familiarity but, as Balota et al. (2001) showed, the former is a better estimate of the relative frequency of exposure to a word than the experiential familiarity construct proposed by Gernsbacher (1984). As was highlighted by Balota et al. (2001), the instructions used by Gernsbacher were extremely vague (Gernsbacher defined a familiar word as a word that participants know and use very often in their everyday life, whereas an unfamiliar word is defined as a word that participants had never seen before and that cannot be recognized) and may have allowed for other word properties (e.g., semantic, orthographic, phonological) to affect ratings. Alternatively, the redefinition of familiarity as the estimation of how often participants come across words in their daily lives (i.e., by explicitly asking participants to rate the number of times they have encountered a word in their written or spoken form) offers a clearer way to assess the relative frequency of exposure to a word than the familiarity measure intends to capture. Indeed, Balota et al. (2004; Balota et al., 2001) and others (e.g., Alderson, 2007; Thompson & Desrochers, 2009) found that the subjective frequency ratings

were highly predictive of both lexical decision and naming performance above and beyond Gernsbacher's (1984) familiarity concept. Hence, since then subjective frequency has become a more suitable way to assess the impact of the subjective exposure to a word on lexical representations and processing, stimulating the collection of subjective frequency norms in different languages.

Indeed, contrary to the words' objective proprieties, mostly obtained from automatic (computational) procedures applied to large corpora (see, e.g., Soares, Iriarte, et al., 2014; Soares, Machado, et al., 2015; Soares, Medeiros, et al., 2014, for recent examples of these procedures), collecting subjective proprieties is more demanding and time-consuming. Typically this implies conducting large-scale studies, and thus asking a great number of participants to rate a set of words in a given subjective dimension. Conducting these studies in the last decades has allowed for the creation of standardized norms for several subjective indices in different languages (e.g., Altarriba et al., 1999; Balota et al., 2001; Barca, Burani, & Arduino, 2002; Bird et al., 2001; Bradley & Lang, 1999; Briesemeister, Kuchinke, & Jacobs, 2011; Brysbaert, Warriner, & Kuperman, 2014; Della Rosa et al., 2010; Desrochers, Liceras, Fernandez-Fuertes, & Thompson, 2010; Ferrand et al., 2008; Gilhooly & Logie, 1980; Kuperman et al., 2012; Paivio et al., 1968; Schmidtke, Schröder, Jacobs, & Conrad, 2014; Stadthagen-Gonzalez & Davis, 2006).

Despite the widespread use of these norms in cognitive research in general and in psycholinguistics in particular, their availability for Portuguese is very limited. Up to now, the few norms available have been the recent AoA norms from the works of Marques and colleagues (Marques, Fonseca, Morais, & Pinto, 2007) and Cameirão and Vicente (2010) for 834 and 1,749 Portuguese words, respectively, but based on different data collection procedures (Marques et al., 2007, collected the AoA ratings using a 7-point scale based on Gilhooly and Logie's, 1980, procedure, whereas Cameirão and Vicente used a 9-point scale following Carroll and White's, 1973, work). Subjective norms for valence, arousal, and dominance were also collected for 1,034 words in the recent adaptation of the Affective Norms for English Words (ANEW; Bradley & Lang, 1999) into European Portuguese (Soares, Comesaña, Pinheiro, Simões, & Frade, 2012). Norms for imageability and concreteness are very scarce—only available for 808 words in Marques et al.'s (2007) work—and for subjective frequency they are nonexistent. For familiarity, norms for a limited pool of Portuguese words (459) are available from Marques et al.'s (2007) database. The lack of subjective norms for Portuguese is thus a major obstacle for conducting cognitive and neuroscientific research using verbal stimuli in Portuguese. Indeed, regardless of the weight these variables might have on the prediction of subjects' performances in psycholinguistic and memory tasks, the literature has shown that they are relevant variables that should not

---

[1] The dual-coding theory makes this prediction because concrete words would benefit from the activation of both verbal and nonverbal (imagistic) codes (Paivio, 1971, 1986), and the context availability theory, because concrete words are strongly linked to a narrow range of supporting contexts in memory, which eases their retrieval, relative to abstract words that are weakly linked to a wide range of contexts (Schwanenflugel, 1991; Schwanenflugel & Shoben, 1983).

be neglected when planning for experimental studies that use verbal stimuli. Disregarding their control may lead to important confounds that can bias the results and threaten the validity of the conclusions. Therefore, having reliable standardized norms for these attributes has become a critical requirement for neuroscientific and cognitive research today.

In this work we aim to overcome this gap by providing subjective norms of imageability, concreteness, and subjective frequency for a large set of Portuguese words (3,800) in the Minho Word Pool (MWP) dataset. It is worth noting that the MWP also integrates words that match those in the above-mentioned national (Cameirão & Vicente, 2010; Marques et al., 2007; Soares et al., 2012) and international databases: namely, the Bristol norms of Stadthagen-Gonzalez and Davis (2006)—one of the biggest subjective databases used in research, which provides AoA, imageability, and familiarity norms for 3,394 English words—the recent work of Brysbaert et al. (2014), which presents norms of concreteness for 40,000 English lemmas, and the Balota et al. (2001) norms, which provide subjective frequency ratings for 2,938 English words. This option allowed us not only to cross-validate the MWP with its national and international counterparts in which the same and other subjective variables (i.e., familiarity, AoA, valence, arousal, and dominance) are available, but additionally to contribute to a more complete characterization of the Portuguese stimuli typically used in experimental research. Thus, future studies can be conducted using verbal stimuli in Portuguese controlled for a large number of subjective indices, which makes the MWP an even more powerful tool for research.

Method

Participants

Two thousand three hundred fifty-seven undergraduate students (1,508 females and 849 males; M_{age} = 22.4, SD = 5.03) from different courses (Humanities, Economics, Sciences, and Technologies) in several public and private universities from the North to the South of Portugal participated in this study. This sample excludes participants whose native language was not European Portuguese or whose nationality was not Portuguese (n = 124), as well as those who did not answer more than 33 % of the items (n = 96), or whose responses demonstrated nondiscriminative ratings and/or random or inattentive responses (e.g., choosing the same number for the majority of the words; n = 88). Thus, from an initial sample of 2,665 participants, 308 participants were excluded from the computation of the normative values of imageability, concreteness, and subjective frequency presented in the MWP. All participants included in the normative study (N = 2,357) were European Portuguese native speakers from all

Portuguese districts, including Madeira and the Azores islands. The majority was right-handed (92.1 %) and had normal (54.6 %) or corrected-to-normal (45.4 %) visual acuity.

Materials

Three thousand and eight hundred words were selected from the Procura PALavras (P-PAL; Soares, Iriarte, et al., 2014) database (available online along with this article or at http://p-pal.di.uminho.pt/tools) to be integrated in the MWP dataset. These words were selected on the basis of the following criteria: (1) if they classified as content words (e.g., nouns, adjectives), since the bulk of the cognitive research using verbal stimuli has focused on this type of words; (2) if they presented different values of occurrence in the language (i.e., different values of lexical frequency), in order to ensure the existence of words from different frequency ranges in the MWP dataset; (3) if they presented different lengths in number of letters and syllables, so that the MWP included both monosyllabic words (the most commonly used in research) and also words of other syllable lengths, in line with the recent claims in the psycholinguistic literature (see, e.g., Yap & Balota, 2009). This option was also justifiable by the fact that Portuguese is a language in which the vast majority of words extend beyond one syllable. For example, in the P-PAL psycholinguistic database mentioned above (which integrates approximately 208,000 word forms), only 641 words are monosyllabic, which corresponds to 0.3 % of the entire lexicon (see Soares, Iriarte, et al., 2014, and Soares, Machado, et al., 2015, for details). Therefore, the MWP was designed to include a more diversified set of words in terms of word length (in number of letters: M = 7.16, SD = 2.08, range: 2 to 12; and in number of syllables: M = 3.14, SD = 0.96, range: 1 to 6) and per million word frequency (M = 39.52; SD = 85.40; range: 0.01 to 1,214.4), which will provide not only a closer representation of the lexical diversity of the Portuguese language, but importantly more versatility in the selection of stimuli, allowing researchers to control for and/or manipulate a series of objective characteristics while also manipulating and/or controlling for subjective characteristics of the words in the MWP.

Figure 1 presents the distribution of the 3,800 words in the MWP as a function of word length in number of letters (i.e., short words, whose length varies between two and five letters; medium words, whose length varies between six and eight letters; and long words, whose length varies between nine and 12 letters) and per-million word frequency (i.e., low-frequency words, ≤10 occurrences per million; medium-frequency words, 11–74 occurrences per million; and high-
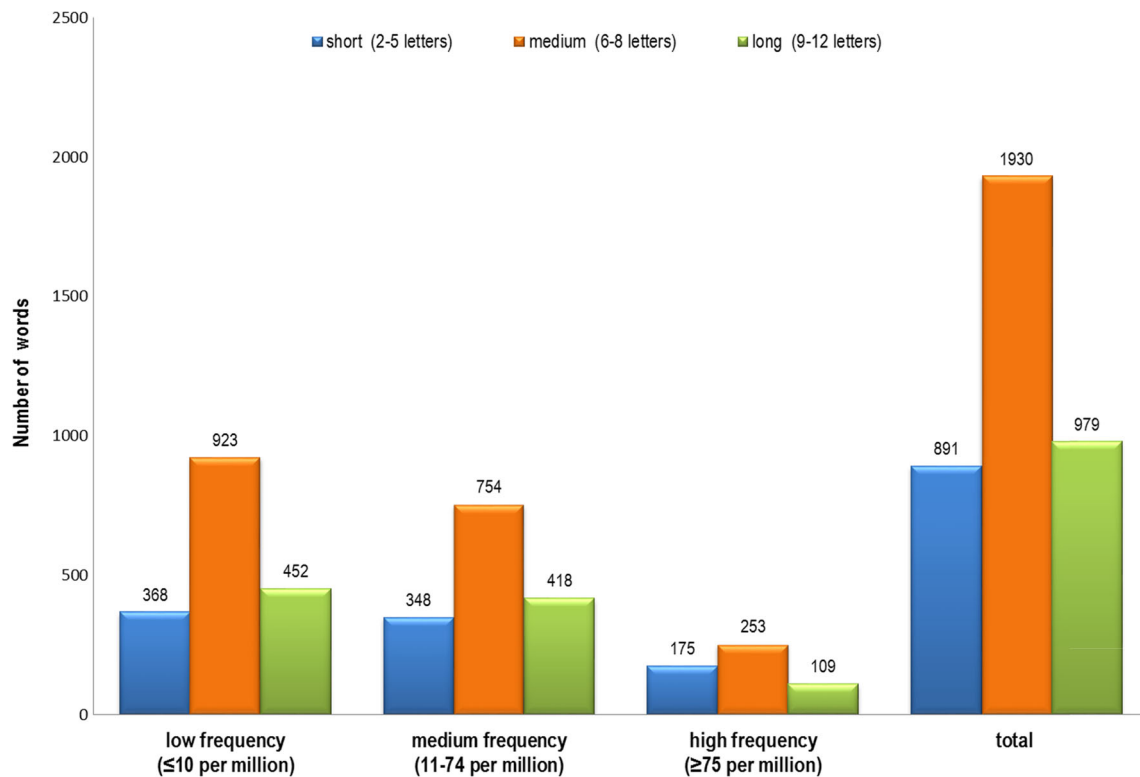
Fig. 1 Distribution of the 3,800 words of the MWP, according to per-million written word-frequency intervals (low, medium, and high) and word length intervals (short, medium, and long words), as obtained from the P-PAL database (Soares, Iriarte, et al., 2014)

frequency words, ≥75 occurrences per million), as obtained from the P-PAL word form database.[2]

As is depicted in Fig. 1, most words in the MWP dataset are medium-length words (50.8 %), followed by long (25.8 %) and short words (23.4 %). Although word length in Fig. 1 was analyzed considering the number of letters, the distribution in the MWP considering the number of syllables showed that three-syllable words are the most frequent in the dataset (39.7 %), followed by two-syllable (25.7 %) and four-syllable (24 %) words. Monosyllables represent only 1.4 % of the entire MWP corpus, and words with more than four syllables represent 9.2 % of the corpus.

Regarding word frequency, most MWP words were low-frequency (45.8 %), followed by medium-frequency (40 %) and high-frequency (14.1 %) words. Including fewer high-frequency words was mainly due to the fact that most high-frequency words in the P-PAL word form database are function words or verb inflections (see Soares, Iriarte, et al., 2014, for details), which were excluded from the MWP dataset. Nevertheless, low-frequency words seem to be more useful

for research, since the majority of psycholinguistic phenomena are observed for low- but not for high-frequency words. It is worth noting that words of different lengths have been incorporated into each word frequency interval (see Fig. 1), although the distribution shows that there are more medium words in each frequency interval than any other word length group (52.9 % of the words in the low-frequency interval, 49.6 % of the words in the medium-frequency interval, and 47.9 % of the words in the high-frequency interval).

Following the suggestion by Stadthagen-Gonzalez and Davis (2006), we also included words whose norms of imageability, concreteness, and subjective frequency were already available from other national (Marques et al., 2007) and international (Balota et al., 2001; Brysbaert et al., 2014; Stadthagen-Gonzalez & Davis, 2006) datasets. Specifically, the MWP contains 221 words whose norms of imageability and concreteness were available for Portuguese from the Marques et al. (2007) norms. Regarding the international databases, the MWP includes 781 words matching the English words in the Bristol norms by Stadthagen-Gonzalez and Davis (2006) in the subjective measures of imageability and familiarity; and 927 words matching those in the subjective frequency norms provided by Balota et al. (2001). Finally, the MWP also includes 3,478 words that match the recent norms for concreteness developed by Brysbaert et al. (2014). This procedure allowed us to cross-validate the MWP dataset with its national and international counterparts by comparing our

---

[2] Although high-frequency words tend to be defined in the literature as words with more than 100 occurrences per million words in a corpus, we established a lower cutoff point for these words in the MWP (≥75 occurrences per million words) because most words in the P-PAL corpus with a frequency equal to or above 100 occurrences per million words were function words or verb forms, which were excluded from the MWP dataset.

ratings with the ones obtained from those databases in the same subjective dimensions. In the same vein, we cross-checked the MWP words with those whose ratings of valence, arousal, and dominance were already available for Portuguese in the recent adaptation of the ANEW database (Bradley & Lang, 1999; Soares et al., 2012; 912 common words),[3] and also with the recent Portuguese norms for AoA by Cameirão and Vicente (2010; 1,265 common words) and Marques et al. (2007; 739 common words). This additional procedure allowed us not only to further cross-validate the MWP dataset by considering other subjective measures for the same words (note that AoA and familiarity norms are also available from the Bristol norms), but importantly, to extend the MWP norms to other subjective variables (e.g., AoA and familiarity) available in national databases, which will contribute to a more complete characterization of these stimuli. Future studies, hence, can rely on the control/manipulation of a much larger number of subjective measures, thus making the MWP a powerful tool for conducting research with Portuguese participants.

Procedure

A Web survey procedure was used to collect the MWP ratings. Online web surveys have been increasingly used in psychological studies (e.g., Balota et al., 2001; Brysbaert et al., 2014; Kuperman et al., 2012), since they allow for easy access to a larger number of participants with clear advantages for data collection. Acknowledging these advantages, we have developed a Web-based application similar to the one used in the Portuguese adaptation of the ANEW, which had proved to be as reliable as the traditional paper-and-pencil procedure (see Soares et al., 2012 for details).

The experiment was advertised by sending an e-mail to the electronic addresses of students attending different courses from public and private universities in Portugal. In this e-mail, information about the aims of the experiment, the research team, and contacts were provided. Participants were also informed about the task requirements and the time needed to complete the survey, as well as about data confidentiality. It is worth noting that this procedure had previously been authorized by the administrations of each institution and that the experiment was conducted with the approval of the Ethics Committee for Human Research of the University of Minho (Braga, Portugal). In addition, in-person contacts were made with the professors of several institutions to ask them to encourage their students to participate in the Web survey.

---

[3] It is worth noting that, despite the fact that the original version of the ANEW (Bradley & Lang, 1999) and the Portuguese adaptation include 1,034 words (see Soares et al., 2012), we decided to exclude 122 of these words from our data collection because they did not occur in the P-PAL database, thus not having a frequency value associated.

After a first e-mail welcoming the students to participate in the study, two reminders were sent: the first one approximately one week after the first e-mail, and the second one two weeks after the first notification. Participation in the online survey started once the students had clicked the hyperlink provided in the e-mail. The general instructions for performing the experiment were displayed. After providing socio-demographic information (e.g., sex, age, e-mail, educational level, native language, etc.), and giving online informed consent, participants were asked to rate a set of 100 words drawn randomly from the full set of 3,800 words, separately for each of the three subjective measures collected in the MWP. We chose to collect data in the three dimensions separately to avoid potential confounds that could arise from the simultaneous rating of different constructs. The order of presentation of the subjective measures was counterbalanced across participants (six possible orders). Participants were randomly assigned to each of the six possible orders, but equivalent numbers of participants rated the words in each order (approximately 400 valid protocols per order).

The instructions corresponded to the Portuguese translation of those used by Paivio et al. (1968), which have been widely used in similar databases (e.g., Bird et al., 2001; Brysbaert et al., 2014; Cortese & Fugett, 2004; Gilhooly & Logie, 1980; Stadthagen-Gonzalez & Davis, 2006) to assess imageability and concreteness. Likewise, for the subjective-frequency index, we used the Portuguese translation of the instructions used by Balota et al. (2001). As in the original instructions, participants were asked to rate the words using a 7-point Likert scale for each subjective dimension. Specifically, for imageability, participants were asked to indicate how easily a word elicited a mental image, by assigning 1 to a *low-imageability word* and 7 to a *high-imageability word*. For concreteness, participants were asked to indicate the extent to which a word referent could be experienced by the senses, by assigning 1 to a *low-concreteness* (or abstract) word, and 7 to a *high-concreteness* word. For subjective frequency, participants were asked to provide an estimate of how often they encountered a word in their everyday life, with 1 assigned to *words that they had never encountered before*, and 7 to *words they encountered several times a day*. For each subjective index, two word examples were offered to anchor the participants' responses to each of the endpoints of the corresponding scale, according to the values available in the Portuguese norms of Marques et al. (2007). Thus, the words *facto* ["fact"] and *lápis* ["pencil"] were used as examples of a low- and a high-imageability word, respectively. The words *democracia* ["democracy"] and *cadeira* ["chair"] were provided as examples of a low- and a high-concreteness word, respectively. Finally, in the subjective frequency scale, the words *bigorna* ["anvil"] and *país* ["country"] were used as examples for words we are unlikely or very likely to encounter in our everyday life, respectively. It is also worth noting that in addition to the original instructions, in our

procedure participants were asked to signal the words they did not know the meanings of. This option aimed at avoiding random responses for unknown words. The instructions used in the present work are presented in Appendix A.

Words were displayed randomly and individually at the center of the computer screen until participants had responded. No time limit was imposed, although participants were instructed to rate the words as quickly as possible. Words were rated by choosing the number that best matched the participant's judgment for a given subjective dimension, or by choosing the "Unknown" key. As soon as the participant had rated a word, the subsequent word appeared, and the previous rating was automatically stored. Continuing to the next word was not possible until a response had been made. It was also not possible to turn back for reratings. Once having rated 100 words in a given subjective dimension, the instructions for rating the next 100 random words in another subjective dimension were displayed, as mentioned. At the end, participants were thanked for their interest and dismissed. The entire procedure lasted about 30 min per participant.

The MWP database

The normative values for imageability (*Imag*), concreteness (*Conc*), and subjective frequency (*Subj_freq*) from the MWP can be downloaded as supplemental materials with this article and from http://p-pal.di.uminho.pt/about/databases. This archive shows the mean values and standard deviations for the 3,800 Portuguese words in the MWP for the three subjective measures. Words are listed alphabetically and have a specific numeric code (from 1 to 3800). Each word is followed by its English translation. It is worth noting that before computing the normative values provided in the MWP, several verifications were carried out to ensure the integrity of the data. Besides admitting only native European Portuguese speakers of Portuguese nationality, we excluded from the analyses any participants who were unfamiliar with more than a third of the words or who assessed more than 66 % of the words with the same value (nondiscriminative and/or random or inattentive responses). If the same person participated more than once, only the data from the first participation were taken into account. Recurring participation was detected by the use of the same login (e-mail address) at different times and by crosschecking the socio-demographic information provided in the registration data. Additionally, for each subjective dimension, ratings from participants indicating that the word was unknown to them were excluded. These cases were very scarce, leading to the exclusion of 0.78 %, 0.88 %, and 0.41 % of the responses for imageability, concreteness, and subjective frequency, respectively. "Unknown" responses were also very rare, and occurred for words like *urze* "heather," *açucena* "white lily," *arenque* "herring," *esturjão* "sturgeon," *furgão* "freight car," *galé* "galley," *gangrena* "gangrene," *lascivo* "lascivious," *moreia* "morey," and *subterfúgio* "subterfuge," which also have a low objective frequency of occurrence in Portuguese. Subsequently, the mean and standard deviation (*SD*) for each MWP word were calculated on the basis of the remaining data ($N = 232{,}852$ responses in the imageability data, $N = 219{,}413$ responses in the concreteness data, and $N = 217{,}646$ responses in the subjective frequency data). Any ratings 2.5 *SD*s below or above the mean of each item were eliminated. The number of outliers in each subjective dimension was also very low, comprising 0.51 % of the data for imageability, 0.75 % of the data for concreteness, and 0.60 % of the data for subjective frequency. After outlier elimination, the means and *SD*s of the 3,800 MWP words were recalculated for each subjective dimension. In sum, the normative values of imageability, concreteness, and subjective frequency in the MWP were based on 230,905 valid ratings for the imageability dimension, 216,557 valid ratings for the concreteness dimension, and 215,351 valid ratings for the subjective frequency dimension. The average numbers of valid responses per word were 60.8 for imageability (range: 35–68), 57.0 for concreteness (range: 32–62), and 56.7 for subjective frequency (range: 42–77).

Besides these normative values (means and *SD*s), the MWP also provides lexical and sublexical measures for the 3,800 words, obtained from the P-PAL (Soares, Iriarte, et al., 2014) database:

Number of letters (P-PAL_N_{lett}): Number of letters for each MWP word. N_{lett} ranges between two ($n = 11$ words, 0.3 % of the MWP corpus) and 12 letters ($n = 84$ words, 2.2 % of the MWP corpus), with an average of 7.16 letters *per* word ($SD = 2.08$).

Number of syllables (P-PAL_N_{syll}): Number of syllables for each MWP word. N_{syll} ranges between one ($n = 54$ words, 1.4 % of the corpus) and six syllables ($n = 11$ words, 0.3 % of the corpus), with an average of 3.14 syllables *per* word ($SD = 0.96$).

Syllabic structure (P-PAL_$Syllabic_{struct}$): Set of consonants (C) and vowels (V) forming the orthographic structure of each MWP word. The database contains 619 different syllable structures. The CV.CV.CV structure, as in *banana* "banana," is the most frequent syllable structure in the MWP ($n = 222$ words, 5.8 % of the corpus), followed by the CV.CV ($n = 207$ words, 5.4 % of the MWP corpus—e.g., the word *bebé* "baby"), the CVC.CV ($n = 191$ words, 5 % of the corpus—e.g., *balde* "bucket"), and the CVC.CV.CV ($n = 149$ words, 3.9 % of the corpus—e.g., *canguru* "kangaroo") syllable structures.

Part of speech (P-PAL_PoS): Morpho-syntactic information for each MWP word as obtained from the P-PAL database. Words in the MWP are content words that cover the following

five grammatical classes: nouns (N), adjectives (ADJ), verbs (V), adverbs (ADV), and interjections (INT), although the majority are nouns (n = 2,616 words, 68.8 % of the corpus) and adjectives (n = 1,159 words, 30.2 % of the MWP corpus). It is worth noting that several categories can co-occur for the same MWP word because syntactic ambiguity is very common in Portuguese (e.g., words like *activo*"active" can occur both as an ADJ or a N in Portuguese). In these cases *P-PAL_PoS* provides all syntactic categories the word can assume, comma-separated.

Objective word frequency (P-PAL_FREQ$_{mil}$): Number of occurrences *per* million words in the P-PAL word form corpus for each MWP word. *P-PAL_FREQ$_{mil}$* ranges between 0.01 (the words *dióspiro* "persimmon" and *térmite* "termite") and 1,214.45 occurrences (the word *ano* "year"). On average the printed objective word frequency of MWP words is 39.52 per million words (SD = 85.40).

Log$_{10}$ objective word frequency (P-PAL_FREQ$_{log10}$): Base-10 logarithm for each MWP word, obtained from the P-PAL word form corpus (computed from P-PAL_FREQmil + 1). *P-PAL_FREQ$_{log10}$* ranges between .004 (the words *dióspiro* "persimmon" and *térmite* "termite") and 3.09 occurrences (the word *ano* "year"), with an average of 1.10 log10 occurrences (SD = 0.66).

Zipf objective word frequency (P-PAL_FREQ$_{Zipf}$): Number of times each MWP word appears in the P-PAL word form corpus in a logarithm 7-point Likert scale as recently proposed by van Heuven, Mandera, Keuleers, and Brysbaert (2014). The Zipf scale is assumed to be an easier way to understand word frequency since word frequency ranges from 1 to 7 points, with the values 1 to 3 indicating low-frequency words and values from 4 to 7 indicating high-frequency words (see van Heuven et al., 2014, for details). In the MWP, *P-PAL_FREQ$_{Zipf}$* ranges from 0.99 (the words *dióspiro* "persimmon" and *térmite* "termite") to 6.08 (the word *ano* "year"), with an average of 3.97 (SD = 0.83).

Orthographic Neighborhood size (P-PAL_ON): Number of orthographic neighbors of each MWP word in the P-PAL word form corpus. *P-PAL_ON* is defined as the number of words of the same length that can be formed by replacing one letter with another, and maintaining the remaining letters constant in the same positions (Coltheart et al., 1977). In the MWP, *P-PAL_ON* ranges from zero (n = 1,003 words, 26.4 % of the MWP corpus) to 27 (n = 1, 0.3 % of the MWP corpus), with an average of 2.92 neighbors *per* word (SD = 4.15).

Orthographic Levenshtein Distance (P-PAL_OLD$_{20}$): Minimum number of operations (i.e., letter substitution, insertion, or deletion) necessary to transform one word into another

considering its 20 closest orthographic neighbors (Yarkoni et al., 2008) in the P-PAL word form corpus. In the MWP, *P-PAL_OLD* ranges from one (n = 142 words, 3.7 % of the MWP corpus) to 5.2 (n = 1 words, 0.3 % of the MWP corpus), with an average of 2.07 operations *per* word (SD = 0.59).

Results and discussion

The results from the normative study of the MWP are presented in two different sections. First, we present the results from the cross-validation of the MWP taking into account the ratings for the same subjective measures (i.e., imageability, concreteness, and subjective frequency), available at other national (Marques et al., 2007) and international reference databases (Balota et al., 2001; Brysbaert et al., 2014; Stadthagen-Gonzalez & Davis 2006). Secondly, we compare the ratings of imageability, concreteness, and subjective frequency from the MWP with the ratings obtained from other subjective measures (e.g., familiarity, AoA, valence, arousal, and dominance) available at those (Balota et al., 2001; Brysbaert et al., 2014; Marques et al., 2007; Stadthagen-Gonzalez & Davis, 2006) as well as two other Portuguese databases (Cameirão & Vicente, 2010; Soares et al., 2012). With this two-step series of analyses, we aimed not only to cross-validate the measures provided in the MWP with its national and international counterparts in which the same subjective measures are available (construct validity), but additionally to contribute to analyzing how the subjective measures provided in the MWP related to the subjective measures available in other national and international databases (convergent/discriminate validity). As we pointed out in the introduction, this will contribute not only to a broader picture of how those theoretical constructs relate to one another empirically (a hotly discussed issue in the current literature) but, importantly, to a more comprehensive characterization of the stimuli provided in the MWP, which would allow for a more appropriate selection of stimuli. The validity and reliability of the MWP norms were tested conducting correlation and internal consistency analyses (Cronbach's alpha), based on the ratings obtained for the same words across languages and databases.

Specifically, for the first block of analyses we considered the ratings of imageability and concreteness for the 221 words available in both the MWP and the Portuguese norms from Marques et al. (2007). For the international databases, we considered imageability ratings for the 781 words in common with the Bristol norms (Stadthagen-Gonzalez & Davis, 2006), ratings of subjective frequency for the 927 words in common with the Balota et al. (2001) norms, and ratings of concreteness for the 3,478 words in common with the recent Brysbaert et al. (2014) norms. Note that although subjective-frequency ratings are absent from Portuguese databases, in the second block of analyses we considered the familiarity ratings

from Marques et al.'s database for the same 714 words, since both constructs are highly correlated. In the same vein, we have also included comparisons to the familiarity ratings from the Bristol norms for the 781 matching words. Moreover, in this second block of analyses, we also integrated other subjective variables that have been shown to account for significant percentages of variance in word recognition and naming latencies (e.g., AoA, word affective content).

AoA ratings were obtained from the Portuguese databases of both Marques et al. (2007) and Cameirão and Vicente (2010)—for totals of 739 and 1,265 matching words, respectively—because each database used different AoA data collection procedures, as we mentioned above. The AoA ratings from the Bristol norms (Stadthagen-Gonzalez & Davis, 2006) were also considered for a pool of 781 matching words. Finally, the affective ratings of valence, arousal, and dominance from the Portuguese adaptation of ANEW (ANEW-PT; Soares et al., 2012) were also included for a pool of 912 matching words.

Table 1 presents descriptive statistics for the subjective indices of imageability, concreteness, and subjective frequency in the MWP, as well as for the indices of imageability, concreteness, familiarity, and AoA for the words also available in other national (Cameirão & Vicente, 2010; Marques et al., 2007) and international (Balota et al., 2001; Brysbaert et al., 2014; Stadthagen-Gonzalez & Davis, 2006) databases. It also presents descriptive statistics for the Portuguese affective norms of valence, arousal, and dominance (Soares et al., 2012) for males and females considered simultaneously (all norms).[4]

Correlations between the MWP and the same subjective psycholinguistic variables

Table 2 presents linear correlations (Pearson) with alpha corrections (Holm) for multiple correlations[5] between the imageability, concreteness, and subjective frequency ratings in the MWP and other national (Marques et al., 2007) and international (Balota et al., 2001; Brysbaert et al., 2014; Stadthagen-Gonzalez & Davis, 2006) databases also containing these subjective variables.

---

[4] Note that although Soares et al. (2012) reported significant differences in the ways that Portuguese females and males rated affective words, thus recommending the use of separate norms for females and males, in this work we chose to use the norms for the entire sample (all norms combined) because the MWP norms were obtained from a combined sample of males and females—that is, both genders considered simultaneously.

[5] Since performing multiple correlations increases the probability of Type I errors (false positives), the critical p values were adjusted using the Holm correction. We chose this method over the Bonferroni correction because the latter dramatically increases the risk of incurring Type II errors (false negatives; see Aickin & Gensler, 1996, for details about the advantages of using the Holm over the Bonferroni method).

As is presented in Table 2, the imageability, concreteness, and subjective frequency ratings in the MWP and other national and international databases under analysis are significantly correlated. Specifically, the imageability ratings in the MWP are strongly correlated with the Portuguese norms by Marques et al. ($r = .92$, $p < .001$) and the English Bristol norms ($r = .77$, $p < .001$). It should be noted that the interrater reliabilities for imageability between the MWP norms and both the Marques et al. ($\alpha = .90$) and Bristol ($\alpha = .85$) norms are also very high, which lends strong support to the idea that, as expected, the imageability ratings in the MWP capture almost the same information as their national and international counterparts.

The results also show that the MWP imageability ratings correlate strongly with the remaining subjective measures depicted in Table 2, particularly with the concreteness ratings from the MWP ($r = .88$, $p < .001$), the Marques et al. ($r = .82$, $p < .001$), and the Brysbaert et al. ($r = .77$, $p < .001$) norms. These results were not unexpected and attest to the strong association between imageability and concreteness in the MWP, as had previously been observed in other studies (e.g., Altarriba et al., 1999; Connell & Lynott, 2012; Gilhooly & Logie, 1980; Marques et al., 2007; Paivio et al., 1968). However, if these two constructs are strongly correlated, they differ in the ways they correlate with some of the other constructs under analysis (e.g., subjective frequency), suggesting that they should not be interpreted as being equivalent, as has been increasingly claimed by several authors in the literature (e.g., Connell & Lynott, 2012; Dellantonio et al., 2014; Kousta et al., 2011). Indeed, although the correlation between the MWP imageability ratings and the MWP subjective frequency ratings did not reach statistical significance (a result also observed when we additionally correlated the imageability ratings from Marques et al. with the subjective frequency ratings of the MWP norms), the correlation between the MWP concreteness ratings and the MWP subjective frequency ratings did ($r = -.09$, $p < .001$), although weakly.

Moreover, concerning the concreteness ratings, the results depicted in Table 2 also show that the MWP concreteness ratings correlated significantly with all subjective variables under analysis, except for the subjective frequency measure from the Balota et al. norms. Specifically, the results from Table 2 show that the MWP concreteness ratings are strongly correlated with the concreteness ratings from both the Marques et al. ($r = .97$, $p < .001$) and the Brysbaert et al. ($r = .86$, $p < .001$) norms, which lends support to the validity of the concreteness ratings in the MWP. It should be noted that, similar to the imageability ratings, the interrater reliabilities between the concreteness ratings in the MWP and both Marques et al. and Brysbaert et al.'s norms were also very high ($\alpha s = .97$ and .92, respectively), providing additional compelling evidence of the reliability of the concreteness ratings in the MWP database.

The MWP concreteness ratings are also strongly correlated with other subjective variables—namely, with the imageability

Table 1 Means (*M*), standard deviations (*SD*), and range values (minimum–maximum) for imageability, concreteness, and subjective frequency from the MWP and for the same and other subjective measures available in national (Cameirão & Vicente, 2010; Marques et al., 2007; Soares et al., 2012) and international (Balota et al., 2001; Brysbaert et al., 2014; Stadthagen-Gonzalez & Davis, 2006) databases, for the common words

Databases	Subjective Dimensions	*N*	*M*	*SD*	Range (min–max)
MWP norms	Imageability (1–7 scale)	3,800	4.50	1.05	2.09–6.75
	Concreteness (1–7 scale)	3,800	4.55	1.30	1.87–6.91
	Subjective frequency (1–7 scale)	3,800	4.20	1.08	1.42–6.93
Marques et al. (2007) norms	Imageability (1–7 scale)	221	4.94	1.64	2.04–6.87
	Concreteness (1–7 scale)	221	4.94	1.93	1.64–6.82
	Familiarity (1–5 scale)	714	2.22	0.73	1.11–4.89
	AoA (1–7 scale)	739	3.35	1.24	1.05–7.24
Cameirão & Vicente (2010) norms	AoA (1–9 scale)	1,265	5.68	1.56	1.33–8.78
Soares et al. (2012) norms	Valence$_{all}$ (1–9 scale)	912	5.02	1.82	1.34–8.46
	Arousal$_{all}$ (1–9 scale)	912	4.86	1.13	1.79–7.77
	Dominance$_{all}$ (1–9 scale)	912	4.96	0.91	1.95–7.47
Bristol norms (Stadthagen-Gonzalez & Davis, 2006)	Imageability (100–700 scale)	871	427.19	138.40	100–668
	Familiarity (100–700 scale)	871	501.11	63.16	358–700
	AoA (years)	871	6.87	1.83	2.10–12.60
	AoA (100–700 scale)	871	398.21	123.64	100–700
Balota et al. (2001) norms	Subjective frequency (1–7 scale)	927	4.47	0.98	1.90–6.72
Brysbaert et al. (2014) norms	Concreteness (1–5 scale)	3,478	3.51	1.10	1.19–5.00

ratings from the Marques et al. and Bristol norms (*r*s = .96 and .75, respectively, *p*s < .001). Thus, in line with the above-mentioned results for imageability, highly concrete words in the MWP were rated as more imageable, considering not only the MWP imageability ratings, but also the imageability ratings from the other national and international databases under analysis. Moreover, the results also showed that the more concrete an MWP word is, the less frequently it is estimated to be used in everyday life. Even though this correlation is far from strong, as mentioned above, the negative correlation between concreteness and subjective frequency in the MWP data was not entirely expected. Indeed, a great number of studies have

shown that concrete words are acquired earlier in life than abstract words (e.g., Barca et al., 2002; Bird et al., 2001; Cameirão & Vicente, 2010; Kuperman et al., 2012; Marques et al., 2007). This led us to expect not only a positive relationship between concreteness and AoA ratings, as will be detailed below (Table 3), but also a positive relationship between concreteness and subjective frequency. However, the results obtained did not confirm this prediction.

Although the positive correlation between concreteness and AoA is consistent with the findings from studies showing that the acquisition of abstract words increases as development unfolds (e.g., Stadthagen-Gonzalez & Davis, 2006),

Table 2 Linear correlations between the MWP imageability, concreteness, and subjective frequency ratings and the ratings obtained for the same subjective indices, available from Portuguese (Marques et al., 2007) and international (Balota et al., 2001; Brysbaert et al., 2014; Stadthagen-Gonzalez & Davis, 2006) databases, for the common words

Databases	Subjective Dimensions	MWP Norms		
		Imageability	Concreteness	Subjective Frequency
MWP norms	Concreteness	.88[**]	–	–
	Subjective frequency	.04[ns]	–.09[**]	–
Marques et al. norms (2007)	Imageability	.92[**]	.96[**]	–.10[ns]
	Concreteness	.82[**]	.97[**]	–.17[ns]
Bristol norms (Stadthagen-Gonzalez & Davis, 2006)	Imageability	.77[**]	.75[**]	–.07[ns]
Balota et al. (2001) norms	Subjective frequency	.01[ns]	–.09[ns]	.71[**]
Brysbaert et al., (2014) norms	Concreteness	.77[**]	.86[**]	–.17[**]

[**] *p* < .001, [*] *p* < .05, [ns] nonsignificant

Table 3 Linear correlations between the MWP imageability, concreteness, and subjective frequency ratings and the ratings obtained for other subjective indices (familiarity, AoA, valence, arousal, and dominance), available from Portuguese (Cameirão & Vicente, 2010; Marques et al., 2007; Soares et al., 2012) and international (Stadthagen-Gonzalez & Davis, 2006) databases, for the common words

Databases	Subjective Dimensions	MWP Norms		
		Imageability	Concreteness	Subjective Frequency
Marques et al. norms (2007)	Familiarity	−.70**	−.58**	−.50**
	AoA	−.66**	−.49**	−.65**
Cameirão & Vicente (2010) norms	AoA	−.62**	−.54**	−.60**
Soares et al. (2012) norms	Valence	.20**	.10 ns	.41**
	Arousal	−.15**	−.25**	−.14**
	Dominance	.10*	.04 ns	.40**
Bristol norms (Stadthagen-Gonzalez & Davis, 2006)	Familiarity	.02 ns	−.07 ns	.62**
	AoA (years)	−.42**	−.35**	−.35**
	AoA (100–700 scale)	−.43**	−.36**	−.36**

** $p < .001$, * $p < .05$, ns nonsignificant

the negative correlation between concreteness and subjective frequency is not immediately understandable, and could be associated with the fact that abstract words tend to be linked to a wider range of contexts in memory than concrete words, as the context availability theory (Schwanenflugel, 1991; Schwanenflugel & Shoben, 1983) and the recent views of embodied cognition (e.g., Barsalou et al., 2008; Dellantonio et al., 2014; Kousta, Vinson, & Vigliocco, 2009, Kousta et al., 2011; Vigliocco et al., 2009, 2013) have claimed. The wide range of associations of abstract words in memory can hence justify the fact that they tend to be used more often in our everyday life than concrete words, which conversely tend to be linked to a more narrow range of associations in memory. This interpretation is also supported by an additional analysis considering the contextual diversity measures obtained from the SUBTLEX-PT database (a new lexical database for Portuguese that provides frequency norms extracted from a subtitle corpus; Soares, Machado, et al., 2015), which indexes the different numbers of contexts in which a word appears ($r = -.06$, $p < .001$)—note that this measure in not available at the P-PAL database.

Finally, concerning the MWP subjective frequency ratings, our results revealed a significant correlation with the subjective frequency norms of Balota et al. ($r = .71$, $p < .001$), as expected. The interrater reliability of subjective frequency in that database was higher ($\alpha = .83$), thus lending strong support to the validity of the subjective frequency norms in the MWP. Besides, the MWP subjective frequency ratings correlated significantly with the concreteness ratings provided by Brysbaert et al.'s norms ($r = -.17$, $p < .001$), but interestingly, not with the concreteness ratings in the MWP. Although the association between the MWP subjective frequency measures and the concreteness ratings from the Brysbaert et al.'s norms is moderate at best, this is an interesting result that seems to

point to important differences in the ways that participants immersed in different languages and cultures rate the degrees to which they use words in their everyday life (subjective frequency) and/or the degrees to which they rate the level of concreteness of words' referents in each language. This provides further evidence of the need to develop standardized norms that respond to those specificities, like the ones presented here for Portuguese. Finally, the absence of any other statistically significant correlation between the subjective frequency ratings in the MWP and the remaining measures in Table 2 seems to reflect that, at least for the Portuguese data, subjective frequency is a construct distinct from any other provided in the MWP.

In the next set of analyses, we explored how the subjective measures in the MWP related to other subjective measures that have been shown to account for significant percentages of variance in word recognition and naming latencies, and that are available in the Portuguese and international databases for the same words.

Correlations between the MWP and other subjective psycholinguistic variables

Table 3 presents the linear correlations (Pearson), with alpha corrections (Holm) for multiple correlations, between the imageability, concreteness, and subjective frequency ratings in the MWP and the ratings of familiarity, AoA, valence, arousal, and dominance from other national (Cameirão & Vicente, 2010; Marques et al., 2007; Soares et al., 2012) and international (Balota et al., 2001; Brysbaert et al., 2014; Stadthagen-Gonzalez & Davis, 2006) databases. As was mentioned above, these supplemental analyses aimed to enrich the cross-validation of the MWP by providing additional evidence for the convergent/discriminant validity of the

constructs presented in this dataset, and also to contribute to a more complete characterization of the stimuli provided in the MWP, which would allow for more appropriate stimulus selection when planning for experimental studies that use verbal materials.

As is depicted in Table 3, the imageability, concreteness, and subjective frequency ratings from the MWP correlated significantly with the other subjective measures provided in the national and international databases under analysis. Specifically, imageability correlated significantly with the familiarity ratings from Marques et al. ($r = -.70$, $p < .001$), though not with the familiarity ratings from the Bristol norms, despite the fact that the instructions used to collect the imageability data in both cases were based on the same instructions, devised by Paivio et al. (1968). Thus, high-imageability words in the MWP were rated not only as more concrete than low-imageability words—as the imageability–concreteness correlation showed in the previous analyses—but also as more familiar, at least as compared to the Portuguese ratings obtained from Marques et al.'s norms (note that the negative correlation between imageability and familiarity is due to the fact that in Marques et al.'s work the familiarity scale was inverted, ranging from 1 *highly familiar* to 5 *very unfamiliar*). The nonsignificant correlation with the familiarity ratings from the Bristol norms might also point to important language/cultural differences in the ways that participants immersed in different languages and cultures rate the ease with which a word evokes a mental image (imageability) and/or the degrees to which they know and use words in their everyday life (familiarity), as was previously mentioned for the absence of statistically significant correlations between the MWP concreteness ratings and the subjective frequency ratings from Balota et al.'s norms.

The MWP imageability ratings also correlated significantly with all AoA ratings presented in Table 3, particularly with the AoA ratings from the national databases by Marques et al. ($r = -.66$, $p < .001$) and Cameirão and Vicente ($r = -.62$, $p < .001$)—the correlations with the two measures of AoA provided by the Bristol norms (AoA in number of years and AoA on a 100–700 scale) were –.42 and –.43 (ps $< .001$), respectively. These findings corroborate previous results (e.g., Barca et al., 2002; Bird et al., 2001; Marques et al., 2007; Stadthagen-Gonzalez & Davis, 2006; Toglia & Battig, 1978) and show that in our data, the more imageable a word is, the earlier in life it tends to be acquired. Finally, the correlations between the imageability ratings in the MWP and the affective ratings from the Portuguese adaptation of the ANEW (Soares et al., 2012) were also statistically significant in all affective dimensions. Specifically, the MWP imageability ratings correlated positively with both valence ($r = .20$, $p < .001$) and dominance ($r = .10$, $p = .029$), and negatively with arousal ($r = -.15$, $p < .001$), although only moderately. These findings show that the ease with which a word evokes a mental image is not dissociated of its emotional content, as has been suggested by several authors (e.g., Altarriba & Bauer, 2004; Dellantonio et al., 2014; Kousta et al., 2011; Paivio et al., 1968; Vigliocco et al., 2013). In the MWP, high-imageability words were rated as more pleasant and less arousing than low-imageability words, and also with higher levels of dominance.

Although a deeper understanding of these results goes beyond the scope of this article, it is important to note that the negative relationship between imageability and arousal was somewhat unexpected. Indeed, if emotional words activate more external and internal sensory information (mainly body-related information) than nonemotional words, as recent embodied accounts of cognition (e.g., Barsalou et al., 2008; Dellantonio et al., 2014; Kousta et al., 2011; Kousta et al., 2009; Vigliocco et al., 2009; Vigliocco et al., 2013) state, we would also expect that the more imageable an emotional word is, the more arousing it would be for individuals. However, the negative correlation observed shows the opposite relationship—that is, the less imageable an MWP word is, the higher the arousal it seems to activate. Note that this negative relationship was also observed when we additionally correlated the Marques et al. imageability ratings with the arousal ratings from the Soares et al. norms ($r = -.44$, $p < .001$), and that these findings are also consistent with the results recently reported by Schmidtke et al. (2014) and Riegel et al. (2015) for German and Polish, respectively. Although this finding should be analyzed further in future studies, it is possible that this negative relationship could be accounted for by the asymmetric distribution of negative and positive words in the high- and low-imageability categories in the MWP. Indeed, if on the one hand we assume that negative words are those rated below 5 (the medium point of the 9-point scale used for the affective ratings) and positive words those with ratings above that value, following Soares et al.'s (2012) suggestion, and on the other hand assume that high-imageability words are those rated above 4 points (the medium point of the 7-point imageability scale used) and low-imageability words those with ratings below that value, we observe that in the MWP, more negative words are classified as low- than as high-imageability words, whereas more positive words are classified as high- than as low-imageability words, $\chi^2(1) = 21.59$, $p < .001$. Thus, considering that negative stimuli are typically rated as significantly more arousing than positive stimuli (e.g., Bradley & Lang, 1999; Briesemeister et al., 2011; Kuperman et al., 2012; Riegel et al., 2015; Schmidtke et al., 2014; Soares et al., 2012, Soares et al. 2013; Soares, Pinheiro, et al., 2015), we can explain this result. In fact, a further partial-correlation analysis showed that the relationship between imageability and arousal was no longer statistically significant ($r = -.05$, $p = .118$) when valence (negative vs. positive) was controlled in the MWP. Nevertheless, the interplay between arousal and imageability should be explored in future studies, since it is a neglected topic in research that could foster a deeper

understanding of the "special" status of emotional words in cognitive processing.

Regarding the MWP concreteness ratings, the results from the correlation analyses also show that concreteness correlated significantly with most subjective variables in Table 3. Specifically, they correlated significantly with the familiarity ratings from Marques et al. ($r = -.58$, $p < .001$), but not with the familiarity ratings from the Bristol norms. Thus, in line with the results observed in the imageability data, highly concrete words were also rated as more familiar than abstract words, at least in the Portuguese data (note that the negative relationship between concreteness and familiarity in Marques et al.'s data is explained by the inversion of the familiarity scale in Marques et al.'s work, as mentioned). Despite potential cultural/language differences in the ways that Portuguese and English participants rated familiarity, the absence of a statistically significant correlation between the MWP concreteness ratings and the familiarity ratings from the Bristol norms might be explained by the fact that the familiarity ratings from the Bristol norms resemble the subjective ratings from the MWP and the Balota et al. norms more closely than the familiarity ratings collected by Marques et al., which were closely related with the experiential familiarly construct of Gernsbacher (1984; for the Bristol norms, participants were asked to provide familiarity ratings using a 7-point scale, with 1 assigned to *words that had never been seen* and 7 to *words that were seen very often, nearly every day*). This interpretation is also supported by the fact that the subjective ratings from the MWP correlated more strongly with the familiarity ratings from the Bristol norms ($r = .62$, $p < .001$) than with the familiarity ratings provided in the Marques et al. norms ($r = -.50$, $p < .001$)—again this negative correlation is explained by the inversion of the familiarity scale in Marques et al.'s work.

The MWP concreteness ratings also correlated significantly with all AoA measures in Table 3 both from the national (Marques et al., $r = -.49$, $p < .001$; Cameirão & Vicente, $r = -.54$, $p < .001$) and international (Bristol norms: AoA in years, $r = -.35$, $p < .001$; AoA on 100–700 scale, $r = -.36$, $p < .001$) databases under analysis, showing that in all cases, the more concrete an MWP word was, the earlier in life it was acquired. This is consistent with previous studies (e.g., Barca et al., 2002; Bird et al., 2001; Cameirão & Vicente, 2010; Kuperman et al., 2012; Marques et al., 2007) showing a negative relationship between the two constructs. Moreover, the MWP concreteness ratings also correlated significantly with the arousal ratings from the Soares et al. norms ($r = -.25$, $p < .001$), mirroring the results previously observed for the imageability–arousal association—note that this negative relationship was also observed when we additionally correlated the concreteness ratings from Marques et al.'s norms with the arousal ratings from Soares et al.'s norms ($r = -.47$, $p < .001$). Thus, similar to the high-imageability MWP words,

the high-concreteness words were rated as less arousing than the abstract words in the MWP, although the association observed between valence and dominance did not reach statistical significance.

The discrepancy in the magnitudes of the correlations between imageability and valence and between concreteness and arousal is not readily interpretable. However, in line with what has been stated previously for the imageability–arousal relationship, and also with the claims of embodied theories of cognition (e.g., Barsalou et al., 2008; Dellantonio et al., 2014; Kousta et al., 2009, Kousta et al. 2011; Vigliocco et al., 2009, 2013), it is possible to anticipate that the less perceptive and experience-based a concept is (i.e., the more abstract a word is), the more internal sensory information it would tend to elicit, which, in turn, could trigger a larger psychophysiological reaction, as captured by the arousal affective dimension. Although this explanation should be explored further, the findings suggest that imageability and concreteness seem to elicit different affective responses in individuals. Whereas imageability seems to be a more "valenced" construct, relevant to define which motivational system (defensive vs. appetitive) would be triggered, concreteness seems to be a more "arousable" construct, with a more prominent role in determining the intensity with which each motivational system will be activated, thus lending additional support to not using these two constructs interchangeably (e.g., Dellantonio et al., 2014; Kousta et al., 2011).

Finally, concerning the MWP subjective frequency ratings, they correlated significantly with all the subjective measures in Table 3. Note that although no subjective frequency norms for Portuguese have been available until now, as mentioned, the correlations between the MWP subjective frequency ratings and the familiarity ratings from the Portuguese Marques et al. norms ($r = -.50$, $p < .001$) and the English Bristol norms ($r = .62$, $p < .001$) show strong relationships, particularly in the case of the Bristol norms. This is a surprising result that can be accounted for by considering that the familiarity ratings from the Bristol norms were collected using Gilhooly and Logie's (1980) procedure, which resembles those used to collect the subjective ratings in the MWP and in Balota et al.'s study more closely than the procedure used in Marques et al.'s work, as we mentioned before. Nonetheless, the fact that a higher correlation was obtained between MWP subjective frequency and the familiarity ratings from Balota et al.'s norms than between the MWP and Marques et al. ratings lends additional support to the validity of the subjective frequency norms provided by the MWP, and suggests that the subjective frequency and familiarity constructs could effectively index different aspects of the relative exposure to a word, as was proposed by Balota et al. (2001).

Moreover, it is also interesting to note that the MWP subjective frequency ratings correlated more strongly with the AoA ratings from both the Marques et al. ($r = -.65$, $p <$

.001) and the Cameirão and Vicente ($r = -.60, p < .001$) norms than with the Portuguese familiarity ratings from Marques et al.'s norms—the correlations between the MWP subjective frequency ratings and the AoA ratings obtained from the Bristol norms were also negative and statistically significant: AoA in years: $r = -.35, p < .001$; AoA on 100–700 scale: $r = -.36, p < .001$. Thus, the MWP words rated with a higher estimation of use in everyday life were not only rated as more familiar than those with lower estimations of use, but seem also to have been primarily acquired earlier in life. These findings with the MWP data lend additional support to the theoretical and empirical distinction between the subjective frequency and experiential familiarity constructs, proposed by Balota et al. (2001; note that if both measures were fair representations of the relative exposure to a word, they should have correlated similarly to AoA). Additionally, they also seem to demonstrate that the subjective estimation of the use of a word is greatly dependent on its AoA, suggesting that subjective frequency is also affected by the cumulative experience with words throughout life, as was suggested by Zevin and Seidenberg (2002)—see, however, Stadthagen-Gonzalez and Davis (2006) for evidence against this account.

Finally, the MWP subjective frequency ratings also correlated significantly with all affective measures from the Soares et al. (2012) norms. Specifically, and consistent with what had been observed before for the imageability and concreteness ratings, MWP subjective frequency correlated positively with valence ($r = .41, p < .001$) and dominance ($r = .40, p < .001$), and negatively with arousal ($r = -.14, p < .001$). Note, however, that the MWP subjective frequency measure correlated more strongly with the affective measures provided by Soares et al.'s norms (particularly in the valence and dominance affective dimensions) than did any other subjective measure from the MWP database. Thus, the higher the estimation of exposure to a word in everyday life, the more positive/pleasant, the less arousing, and the higher the dominance ratings it tended to receive. This is an interesting result that seems to show that as the exposure to a word increases, the probability of a positive reaction to it also increases, as the mere-exposure effect (Zajonc, 1968) predicts. These findings also seem to confirm in the Portuguese data the "Polyanna hypothesis," as it was originally termed by Boucher and Osgood (1969), which claims that individuals tend to use words that make them feel happy, more relaxed, and in control more often than equally familiar negative words. This positive bias, found in English and in several other languages (see, e.g., Augustine, Mehl, & Larsen, 2011, and also Dodds et al., 2015, for a recent work with ten different languages) is assumed to reflect a universal tendency for pro-social communication, and thus can account for the fact that the positive words in the MWP were indeed estimated as being used more often than less positive words, which can also contribute to explaining the attentional bias for positive emotional stimuli in the literature

(see Pool, Brosch, Delplanque, & Sander, 2016, for a recent meta-analysis).

Conclusion

In the present study, we have presented the MWP, a database that provides normative values of imageability, concreteness, and subjective frequency for 3,800 (European) Portuguese words. The MWP was developed to respond to the lack of normative values in Portuguese for three of the most widely used subjective indices in the literature, and thus to support cognitive and neuroscientific research using verbal stimuli with Portuguese participants. The 3,800 words selected for the MWP were content words with different ranges of word length and word frequency from the P-PAL psycholinguistic database. The idea was to provide researchers with a diverse set of words that not only more closely represent the lexical diversity of the Portuguese language, but also would enable researchers to control and/or manipulate a series of objective measures while manipulating (or controlling for) the subjective measures available in the MWP. It is worth noting that the MWP also includes words that match those existing in other national (Cameirão & Vicente, 2010; Marques et al., 2007; Soares et al., 2012) and international (Balota et al., 2001; Brysbaert et al., 2014; Stadthagen-Gonzalez & Davis, 2006) databases. This fact allowed us not only to cross-validate the MWP with its national and international counterparts in which these, as well as other, subjective variables are available, but additionally to contribute to a more complete characterization of those stimuli. Thus, future studies with Portuguese verbal stimuli can now be conducted with control and/or manipulation of a broader range of subjective measures, which makes the MWP an even more powerful instrument for research. The MWP norms can be downloaded along with this article and from http://p-pal.di.uminho.pt/about/databases.

Appendix A

Verbatim instructions for the imageability ratings

As palavras diferem na sua capacidade para evocar imagens mentais de coisas ou situações. Enquanto algumas palavras evocam experiências sensoriais muito fácil e rapidamente,

outras conseguem-no com alguma dificuldade e só após um intervalo de tempo alargado. Outras ainda não chegam a evocar qualquer experiência sensorial.

Neste estudo pretendemos que avalie a facilidade com que as palavras evocam imagens mentais. Se uma dada palavra evoca imagens visuais, sonoras, olfativas ou de outro tipo com facilidade e rapidez, deverá ser avaliada com um valor elevado de imaginabilidade. As palavras que evocam imagens mentais com dificuldade e lentamente, ou que não evocam quaisquer imagens mentais, deverão receber uma avaliação baixa de imaginabilidade. Pense, por exemplo, nas palavras "*lápis*" e "*facto*." "*Lápis*" evoca provavelmente uma imagem mental muito fácil e rapidamente, pelo que deverá atribuir-lhe um valor elevado de imaginabilidade. "*Facto*," pelo contrário, evoca muito provavelmente uma imagem mental com maior dificuldade e lentidão e por isso deverá receber uma avaliação baixa de imaginabilidade.

Repare que a maioria das palavras tende a fazer-nos lembrar outras palavras por associação, tal como "*lápis*"–"*canet*a." É importante que considere nas suas avaliações apenas a facilidade com que a palavra apresentada evoca uma imagem mental e não o número de imagens que a palavra evoca. As avaliações deverão ser realizadas numa escala de sete pontos, em que 1 corresponde ao valor de menor imaginabilidade e 7 ao valor mais elevado de imaginabilidade. Entre estes dois extremos encontram-se valores intermédios que poderão representar melhor o grau de imaginabilidade de outras palavras.

baixa imaginabilidade ← ① ② ③ ④ ⑤ ⑥ ⑦ → elevada imaginabilidade

Não se preocupe se usar mais um determinado valor da escala do que outros desde que isso corresponda à sua apreciação verdadeira. Seja rápido nas suas avaliações mas procure ser o mais preciso possível. Se desconhecer alguma palavra assinale-o, por favor, no espaço disponibilizado para esse efeito.

Verbatim instructions for the concreteness ratings

As palavras diferem no grau em que se referem a conceitos concretos ou abstratos. Enquanto algumas palavras são consideradas concretas porque se referem a coisas, objetos, pessoas ou lugares que podem ser vistos, ouvidos, sentidos, cheirados ou saboreados, outras são abstratas porque se referem a conceitos que dificilmente são materializáveis ou experienciados pelos sentidos.

Neste estudo pretendemos que avalie o grau de concreteza ou de abstração de um conjunto de palavras. As palavras que se referem a coisas, objetos, materiais ou pessoas deverão receber uma avaliação elevada de concreteza, enquanto que as palavras que se referem a conceitos abstratos que não

podem ser experimentados pelos sentidos deverão receber uma avaliação elevada de abstração.

Pense, por exemplo, nas palavras "cadeira" e "democracia." A palavra "cadeira" é concreta porque corresponde a um objeto facilmente percecionado pelos sentidos, pelo que deverá receber uma avaliação elevada de concreteza. "*Democracia*," pelo contrário, corresponde a uma palavra abstrata porque se refere a um conceito imaterial e difícil de experimentar pelos sentidos, devendo por isso receber uma avaliação elevada de abstração. As avaliações deverão ser realizadas numa escala de sete pontos, em que 1 corresponde ao valor mais elevado de abstração e 7 ao valor mais elevado de concreteza. Entre estes dois extremos encontram-se valores intermédios que poderão representar melhor o grau de abstração-concreteza de outras palavras.

baixa concreteza ← ① ② ③ ④ ⑤ ⑥ ⑦ → elevada concreteza

Não se preocupe se usar mais um determinado valor da escala do que outros desde que isso corresponda à sua apreciação verdadeira. Seja rápido nas suas avaliações mas procure ser o mais preciso possível. Se desconhecer alguma palavra assinale-o, por favor, no espaço disponibilizado para esse efeito.

Verbatim instructions for the subjective frequency ratings

O número de vezes com que somos confrontados com uma dada palavra no nosso dia-a-dia varia de palavra para palavra. Apesar de encontrarmos ou usarmos algumas palavras muitas vezes, há outras que só aparecem muito raramente.

Neste estudo pretendemos que avalie a frequência com que considera que entra em contacto com uma dada palavra, isto é, o número aproximado de vezes com que as lê, usa ou ouve no seu dia-a-dia. Uma palavra que seja usada ou encontrada muitas vezes deverá ser avaliada com um valor elevado de frequência, enquanto que as palavras que encontra mais raramente, ou que nunca tenha encontrado deverão receber uma avaliação baixa de frequência. Pense, por exemplo, nas palavras "*país*" e "*bigorna*." Certamente que encontra ou usa a palavra "*país*" várias vezes ao dia pelo que deverá atribuir-lhe uma avaliação alta de frequência. "*Bigorna*," pelo contrário, corresponde seguramente a uma palavra muito menos frequente que poderá até nunca ter encontrado ou usado ao longo da sua vida, devendo por isso atribuir-lhe uma avaliação baixa de frequência.

As avaliações deverão ser realizadas numa escala de sete pontos, em que 1 corresponde a palavras que nunca tenha usado ou encontrado e 7 a palavras que usa ou encontra várias vezes ao dia. Entre estes dois extremos encontram-se outros valores que podem representar melhor a sua estimativa quanto ao número aproximado de vezes com que entra em

contacto com uma dada palavra.

① ② ③ ④ ⑤ ⑥ ⑦

nunca   *uma vez por*   *uma vez por*   *uma vez por*   *de dois em*   *uma vez por*   *várias vezes ao*
    ano   *mês*   *semana*   *dois dias*   *dia*   *dia*

Não se preocupe se usar mais um determinado valor da escala do que outros desde que isso corresponda à sua apreciação verdadeira. Seja rápido nas suas avaliações mas procure ser o mais preciso possível. Se desconhecer alguma palavra assinale-o, por favor, no espaço disponibilizado para esse efeito.

References

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word naming and lexical decision times. *Psychological Science, 17,* 814–823.

Aickin, M., & Gensler, H. (1996). Adjusting for multiple testing when reporting research results: The Bonferroni vs. Holm methods. *American Journal Public Health, 86,* 726–728.

Alderson, J. C. (2007). Judging the frequency of English words. *Applied Linguistics, 28,* 383–409.

Altarriba, J., & Bauer, L. M. (2004). The distinctiveness of emotion concepts: A comparison between emotion, abstract, and concrete words. *American Journal of Psychology, 117,* 389–410.

Altarriba, J., Bauer, L. M., & Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods, Instruments, & Computers, 31,* 578–602. doi:10.3758/BF03200738

Augustine, A. A., Mehl, M. R., & Larsen, R. J. (2011). A positivity bias in written and spoken English and its moderation by personality and gender. *Social Psychological and Personality Science, 2,* 508–515.

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

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition, 29,* 639–647. doi:10.3758/BF03200465

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

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. De Vega, A. M. Glenberg, & A. C. Graesser (Eds.), *Symbols and embodiment: Debates on meaning and cognition* (pp. 245–283). Oxford, UK: Oxford University Press.

Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers, 33,* 73–79. doi:10.3758/BF03195349

Boucher, J., & Osgood, C. (1969). The Pollyanna hypothesis. *Journal of Verbal Learning and Behavior, 8,* 1–8. doi:10.1016/S0022-5371(69)80002-2

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings (Technical Report C-1)*. Gainesville, FL: University of Florida, NIMH Center for Research in Psychophysiology.

Briesemeister, B. B., Kuchinke, L., & Jacobs, A. M. (2011). Discrete emotion norms for nouns: Berlin Affective Word List (DENN-BAWL). *Behavior Research Methods, 43,* 441–448. doi:10.3758/s13428-011-0059-y

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology, 58,* 412–424. doi:10.1027/1618-3169/a000123

Brysbaert, M., & Cortese, M. J. (2011). Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *Quarterly Journal of Experimental Psychology, 64,* 545–559. doi:10.1080/17470218.2010.503374

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46,* 904–911. doi:10.3758/s13428-013-0403-5

Cameirão, M. L., & Vicente, S. G. (2010). Age-of-acquisition norms for a set of 1,749 Portuguese words. *Behavior Research Methods, 42,* 474–480. doi:10.3758/BRM.42.2.474

Carroll, J., & White, M. (1973). Age-of-acquisition norms for 220 picturable nouns. *Journal of Verbal Learning and Verbal Behavior, 12,* 563–576. doi:10.1016/S0022-5371(73)80036-2

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.

Connell, L., & Lynott, D. (2012). Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition, 125,* 452–465. doi:10.1016/j.cognition.2012.07.010

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers, 36,* 384–387. doi:10.3758/BF03195585

Della Rosa, P. A., Catricalà, E., Vigliocco, G., & Cappa, S. F. (2010). Beyond the abstract–concrete dichotomy: Mode of acquisition, concreteness, imageability, familiarity, age of acquisition, context availability, and abstractness norms for a set of 417 Italian words. *Behavior Research Methods, 2,* 1042–1048. doi:10.3758/BRM.42.4.1042

Dellantonio, S., Mullatti, C., Pastore, L., & Job, R. (2014). Measuring inconsistencies can lead you forward: Imageability and the x-ception theory. *Frontiers in Psychology, 5,* 798. doi:10.3389/fpsyg.2014.00708

Desrochers, A., Liceras, J. M., Fernández-Fuertes, R., & Thompson, G. L. (2010). Subjective frequency norms for 330 Spanish simple and compound words. *Behavior Research Methods, 42,* 109–117. doi:10.3758/BRM.42.1.109

Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R.···& Danforth, C. M. (2015). Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences, 112,* 2389–2394.

Ferrand, L., Bonin, P., Méot, A., Augustinova, M., New, B., Pallier, C., & Brysbaert, M. (2008). Age-of-acquisition and subjective frequency estimates for all generally known monosyllabic French words and their relation with other psycholinguistic variables. *Behavior Research Methods, 40,* 1049–1054. doi:10.3758/BRM.40.4.1049

Ferrand, L., Brysbaert, M., Keuleers, E., New, B., Bonin, P., Méot, A., ···Pallier, C. (2011). Comparing word processing times in naming, lexical decision, and progressive demasking: Evidence from Chronolex. *Frontiers in Psychology*, *2*, 306:1–10. doi:10.3389/fpsyg.2011.00306

Fliessbach, K., Weis, S., Klaver, P., Elger, C. E., & Weber, B. (2006). The effect of word concreteness in recognition memory. *NeuroImage, 32,* 1413–1421.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and

polysemy. *Journal of Experimental Psychology: General, 113,* 256–281. doi:10.1037/0096-3445.113.2.256

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation, 12,* 395–427. doi:10.3758/BF03201693

Gordon, B. (1985). Subjective frequency and the lexical decision latency function: Implications for mechanisms of lexical access. *Journal of Memory and Language, 24,* 631–645.

Juhasz, B. J., Yap, M. J., Dicke, J., Taylor, S. C., & Gullick, M. M. (2011). Tangible words are recognized faster: The grounding of meaning in sensory and perceptual systems. *Quarterly Journal of Experimental Psychology, 64,* 1683–1691. doi:10.1080/17470218.2011.605150

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General, 140,* 14–34. doi:10.1037/a0021446

Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition, 112,* 473–481. doi:10.1016/j.cognition.2009.06.007

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44,* 978–990. doi:10.3758/s13428-012-0210-4

Marques, J. F., Fonseca, F. L., Morais, A. S., & Pinto, I. A. (2007). Estimated age of acquisition norms for 834 Portuguese nouns and their relation with other psycholinguistic variables. *Behavior Research Methods, 39,* 439–444. doi:10.3758/BF03193013

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English lexicon project. *Psychonomic Bulletin & Review, 13,* 45–52. doi:10.3758/BF03193811

Osgood, C. E., Suci, G., & Tannenbaum, P. (1957). *The measurement of meaning.* Urbana, IL: University of Illinois Press.

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

Paivio, A. (1986). *Mental representations: A dual coding approach.* Oxford, UK: Oxford University Press.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76*(1, Pt. 2), 1–25. doi:10.1037/h0025327

Parmentier, F. B. R., Comesaña, M., & Soares, A. P. (2016). Disentangling the effects of word frequency and contextual diversity on serial recall performance. *Quarterly Journal of Experimental Psychology.* doi:10.1080/17470218.2015.1105268

Perea, M., Soares, A. P., & Comesaña, M. (2013). Contextual diversity is a main determinant of word-identification times in young readers. *Journal of Experimental Child Psychology, 116,* 37–44.

Pool, E., Brosch, T., Delplanque, S., & Sander, D. (2016). Attentional bias for positive emotional stimuli: A meta-analytic investigation. *Psychological Bulletin, 142,* 79–106. doi:10.1037/bul0000026

Riegel, M., Wierzba, M., Wypych, M., Żurawski, Ł., Jednoróg, K., Grabowska, A., & Marchewka, A. (2015). Nencki Affective Word List (NAWL): The cultural adaptation of the Berlin Affective Word List–Reloaded (BAWL-R) for Polish. *Behavior Research Methods, 47,* 1222–1236. doi:10.3758/s13428-014-0552-1

Schmidtke, D. S., Schröder, T., Jacobs, A. M., & Conrad, M. (2014). ANGST: Affective norms for German sentiment terms, derived from the Affective Norms for English Words. *Behavior Research Methods, 46,* 1108–1118. doi:10.3758/s13428-013-0426-y

Schwanenflugel, P. J. (1991). Why are abstract concepts hard to understand? In P. J. Schwanenflugel (Ed.), *The psychology of word meaning* (pp. 223–250). Hillsdale, NJ: Erlbaum.

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials.

Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 82–102. doi:10.1037/0278-7393.9.1.82

Soares, A. P., Comesaña, M., Pinheiro, A. P., Simões, A., & Frade, C. S. (2012). The adaptation of the Affective Norms for English Words (ANEW) for European Portuguese. *Behavior Research Methods, 44,* 256–269. doi:10.3758/s13428-011-0131-7

Soares, A. P., Iriarte, A., Almeida, J. J., Simões, A., Costa, A., França, P., ··· Comesaña, M. (2014). Procura-PALavras (P-PAL): Uma nova medida de frequência lexical do Português Europeu contemporâneo "Procura-PALavras (P-PAL): A new measure of word frequency for contemporary European Portuguese". *Psicologia: Reflexão e Crítica, 27,* 1–14

Soares, A. P., Machado, J., Costa, A., Iriarte, A., Simões, A., Almeida, J. J.···Perea, M. (2015). On the advantages of frequency measures extracted from subtitles: The case of Portuguese. *Quarterly Journal of Experimental Psychology, 68,* 680–696. doi:10.1080/17470218.2014.964271

Soares, A. P., Medeiros, J. C., Simões, A., Machado, J., Costa, A., Iriarte, A.,···Comesaña, M. (2014). ESCOLEX: A grade-level lexical database from European Portuguese elementary to middle school textbooks. *Behavior Research Methods, 46,* 240–253. doi:10.3758/s13428-013-0350-1

Soares, A. P., Pinheiro, A. P., Costa, A., Frade, C. S., Comesaña, M., & Pureza, R. (2013). Affective auditory stimuli: Adaptation of the International Affective Digitized Sounds (IADS-2) for European Portuguese. *Behavior Research Methods, 45,* 1168–1181. doi:10.3758/s13428-012-0310-1

Soares, A. P., Pinheiro, A. P., Costa, A., Frade, C. S., Comesaña, M., & Pureza, R. (2015). Adaptation of the International Affective Picture System (IAPS) for European Portuguese. *Behavior Research Methods, 47,* 1159–1177. doi:10.3758/s13428-014-0535-2

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods, 38,* 598–605. doi:10.3758/BF03193891

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

Thompson, G. L., & Desrochers, A. (2009). Corroborating biased indicators: Global and local agreement among objective and subjective estimates of printed word frequency. *Behavior Research Methods, 41,* 452–471. doi:10.3758/BRM.41.2.452

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

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67,* 1176–1190. doi:10.1080/17470218.2013.850521

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition, 1,* 219–247. doi:10.1515/LANGCOG.2009.011

Vigliocco, G., Vinson, D., Della Rosa, P., Kousta, S. T., Devlin, J., & Cappa, S. F. (2013). The neural representation of abstract words: The role of emotion. *Cerebral Cortex, 24,* 1767–1777.

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

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

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology Monograph Supplement, 9,* 1–27.

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory & Language, 47,* 1–29. doi:10.1006/jmla.2001.2834