Imageability ratings across languages

Adrià Rofes[1], Lilla Zakariás[2], Klaudia Ceder[3], Marianne Lind[4,5], Monica Blom Johansson[3], Vânia de Aguiar[1], Jovana Bjekić[6], Valantis Fyndanis[4], Anna Gavarró[7], Hanne Gram Simonsen[4], Carlos Hernández Sacristán[8], Maria Kambanaros[9], Jelena Kuvač Kraljević[10], Silvia Martínez-Ferreiro[11], İlknur Mavis[12], Carolina Méndez Orellana[13], Ingrid Sör[3], Ágnes Lukács[14], Müge Tunçer[12], Jasmina Vuksanović[6], Amaia Munarriz Ibarrola[15], Marie Pourquie[16], Spyridoula Varlokosta[17], David Howard[18]

[1]Trinity College Dublin, Ireland; [2]University of Potsdam, Germany; [3]Uppsala University, Sweden; [4]University of Oslo, Norway; [5]Statped, Norway; [6]University of Belgrade, Serbia; [7]Universitat Autònoma de Barcelona, Spain; [8]Universitat de València, Spain; [9]Cyprus University of Technology, Cyprus; [10]University of Zagreb, Croatia; [11] University of Copenhagen, Denmark; [12]Anadolu University, Turkey; [13]Universidad Católica de Chile, Chile; [14]Budapest University of Technology and Economics, Hungary; [15]University of the Basque Country, UPV/EHU, Spain; [16]Basque Center on Cognition Brain and Language, Spain; [17]National and Kapodistrian University of Athens, Greece; [18]Newcastle University, UK

Acknowledgements

The Collaboration of Aphasia Trialists (CATs, research network) is funded by the European Cooperation in Science and Technology (COST, IS1208). For more information please visit www.aphasiatrials.org. This project was partially supported by the Global Brain Health Institute (Rofes). The Basque team (Munarriz and Pourquie) was partially supported by the Basque Government (IT983-16 – GIC 15/129) and MINECO/FEDER (FFI2015-68589-C2-1-P). The work by the authors from Norway (Lind and Simonsen) was partly supported by the Research Council of Norway through its Centres of Excellence funding scheme (223265). The Croatian study (Kuvač Kraljević) was supported by the Croatian Science Foundation and the project Adult language processing (ALP, HRZZ-2421-UIP-11-2013). The Catalan study (Gavarró) was supported by project FFI2014-56968-C4-1-P. The Serbian study (Bjekić and Vuksanović) was supported by the Ministry of Education Science and Technological development grant (#IO175012). The Turkish study (Mavis and Tunçer) was supported by Anadolu University, Scientific Research Project (BAP) grant 1509S632. The Spanish study (Martínez-Ferreiro) was partly supported by PROGRAM (University of Copenhagen Excellence Programme for Interdisciplinary Research) and projects from the Ministerio de Economía y Competitividad (FFI2015-68589-C2-1-P and FFI2014-61888-EXP).

Abstract

Introduction: Imageability is a psycholinguistic variable that indicates how well a word gives rise to a mental image or sensory experience. Imageability ratings are used extensively in psycholinguistic, neuropsychological and aphasiological studies. However, little formal knowledge exists on whether and how these ratings are associated between and within languages.

Methods and results: Fifteen imageability databases were cross-correlated using non-parametric statistics. Some of these corresponded to unpublished data collected within a European research network – the Collaboration of Aphasia Trialists (COST IS1208). All but four correlations were significant. The average strength of the correlations (rho =.68) and the variance explained (R^2 =46%) were moderate. This implies that factors other than imageability may explain 54% of the results.

Conclusion: Imageability ratings often correlate across languages. Different possibly interacting factors may explain the moderate strength and variance in the correlations: (1) linguistic and cultural factors; (2) intrinsic differences between databases; (3) range effects; (4) small numbers of words in each database, equivalent words, and participants; and (5) mean age of participants. The results suggest that imageability ratings may be used cross-linguistically. However, further understanding of the factors explaining the variance in the correlations is needed, before research and practice recommendations can be made.

Keywords: Imageability, linguistics, cross-linguistic, correlations

Imageability ratings across languages

Imageability (also named imagery) is a psycholinguistic variable that is used to indicate how well a word gives rise to a mental image or sensory experience. Imageability ratings are typically collected through paper or web-based questionnaires. Words like "apple" or "house", for example, are typically rated as high in imageability, while words such as "fact" or "hope" are rated as low in imageability (Paivio, Yullie, & Madigan, 1968). Imageability ratings are used in empirical studies of language. Examples of these are the association and analogy work of Francis Galton (1822-1911) and Carl Gustav Jung (1875-1961), and the statistical approach of Friedrich Wilhelm Kaeding (1843-1928) and George Kingsley Zipf (1902-1950), among many others (Levelt, 2014: 449). Imageability ratings are also relevant in neuropsychological and aphasiological studies. Published datasets varying in length exist for languages such as Chinese (Ma, Golinkoff, Hirsh-Pasek, McDonough, & Tardif, 2009), English (e.g., Bird, Franklin, & Howard, 2001; Coltheart, 1981; Cortese & Fugett, 2004; Schock, Cortese & Khanna, 2012), French (Desrochers & Thompson, 2009), Italian (Della Rosa, Catricalà, Vigliocco, and Cappa, 2010; Rofes, de Aguiar, & Miceli, 2015), Japanese (Nishimoto, Ueda, Miyawaki, Une, and Takahashi, 2012), Norwegian (Simonsen, Lind, Hansen, Holm, & Mevik, 2013; Lind, Simonsen, Hansen, Holm, & Mevik, 2015), and Swedish (Blomberg & Öberg, 2015). However, despite some of this excellent work, little is known about the association of imageability ratings between and within languages (see Blomberg & Öberg, 2015, for a recent analysis on Swedish and English).

Psycholinguistic studies

In a seminal study, Paivio et al. (1968) found a high positive correlation between imageability and concreteness ratings. The authors stressed that these two variables are not the same, as concreteness ratings have a dichotomous nature, while imageability ratings respond to a scale. For example, the word "apple" is concrete because it refers to an object or material, while "fact" is not concrete, because it cannot be experienced by the senses. At the same time, "apple" is higher in imageability than "fact", but "apple" is also higher in imageability than "appliance" – even though the latter also refers to a concrete object. Paivio et al. (1968) also found that words "associated with sensory experience (usually affective in nature) but not [referring to] specific things or classes of things" such as "affection", "blessing", "ghost", "delirium" and "hierarchy" were higher in imageability than concreteness; while words that had an "infrequent association with [a] concrete sensory experience" such as "antitoxin", "encephalon" and "originator" were higher in concreteness than imageability. Despite those arguments, many scholars have interchangeably used the terms imageability and concreteness (e.g., McMullen & Bryden, 1987; Tyler & Moss, 1997; Tyler, Moss, Galpin, & Voice, 2002). Indeed, there is a high degree of correlation and similarities between the two variables (e.g., in the Medical Research Council [MRC] psycholinguistic database (Coltheart, 1981), the two variables correlate at rho =.84).

Psycholinguistic studies have shown that words that are rated high in imageability are processed differently – typically, faster and more accurately – than low imageability words. This phenomenon has been named "imageability effect" and has been attributed to different factors: from word differences in age of acquisition (e.g., Carrol & White, 1973; Morrison & Ellis, 1995; Stoke, 1929) or perceptual features (e.g., Plaut & Shallice, 1993), to a separate conceptualization of high and low imageability words in the mental lexicon (e.g., Paivio, 2014). Imageability effects have been shown in lexical decision tasks (e.g., Cortese & Schock, 2013; Schwanenflugel, Harnishfeger, & Stowe, 1988; cf. Tyler et al., 2002); in word production paradigms (e.g., Alario, Ferrand, Laganaro, New, Frauenfelder, & Segui, 2004; Strain, Patterson, & Seidenberg, 1995; cf. Bleasdale & Fraser, 1987; Coltheart, Laxton, & Keating, 1988); and word recognition memory (Cortese, Khanna, & Hacker, 2010; Cortese, McCarthy, & Schock, 2015). Imageability effects have also been shown in tasks that use sentences. Holmes and Langford (1976), for example, indicated that healthy individuals recalled less accurately sentences constructed with low imageability words (e.g., "Many factors affected the crucial choice") as opposed to sentences constructed with high imageability words (e.g., "Many sailors deserted the sinking vessel"). Furthermore, neuroimaging studies have indicated an asymmetrical engagement of the left and right perisylvian and entorhinal cortices, when healthy individuals hear or read high imageability words as opposed to low imageability words (e.g., Wise et al., 2000), and also when performing semantic similarity judgment tasks (e.g., Sabsevitz, Medler, Seidenberg, & Binder, 2005).

Imageability ratings have also been used to control experimental conditions in multiple empirical language studies, as failing to do so may create undesired artifactual results. Naming and reading differences have been shown to disappear when items (i.e., nouns vs. verbs; function vs. content words) were matched for imageability in studies on healthy individuals (Davelaar & Besner, 1988), people with dyslexia (Allport & Funnell, 1981), and people with aphasia (e.g., Hanley & Kay, 1997; Howard & Franklin, 1988; Franklin, Howard, & Patterson, 1995). These results are in contrast with other studies on people with aphasia where, even when matching items for imageability, differences were found in naming and sentence completion tasks (e.g., Berndt, Haendiges, Burton & Mitchum, 2002; Kambanaros & Grohmann, 2015; Rofes, Capasso, & Miceli, 2015). In relation to this, it has been argued that even when nouns and verbs are matched for imageability, there may still exist differences in the cognitive processes necessary to process nouns and verbs, as participants take significantly longer to rate the imageability of nouns than verbs (Chiarello, Shears, & Lund, 1999). Other work where experimental stimuli were matched for imageability include studies finding separate effects of imageability and grammatical class during single-word comprehension using fMRI (Bedny & Thompson-Schill, 2006); studies testing the efficacy of a linguistically motivated protocol to treat people with post-stroke aphasia (de Aguiar et al., 2015); cross-linguistic comparisons of bilectal children speaking Greek and Cypriot Greek (Kambanaros, Grohman, & Michaelides, 2013); and effects of context and word class on the retrieval of words in Chinese speakers with aphasia (Law, Kong, Lai, & Lai, 2015). Further discussion over different ways of matching items is beyond the scope of this paper.

Neuropsychological and aphasiological studies

Imageability ratings along with ratings for frequency, word length, regularity of spelling, and grammatical category are considered a source of evidence to identify impairments to specific levels of language processing. Other relevant sources of information include number and type of errors. Shallice (1988) called this the "critical variable approach". This approach has helped us to understand the underlying deficits that explain why a person with deep dyslexia may read "sandal" when given the word "scandal". In this example, it is assumed that reading "scandal" may also activate the word representation of "sandal", as the words are very similar at the orthographic level, along with other words. The production of "sandal" will be favored over "scandal" if the person has an impairment in abstract word semantics (where imageability plays an important role). This is because the word "sandal" has a higher imageability value than "scandal" (see, e.g., Whitworth, Webster, & Howard, 2014, p.11).

In other studies, people with aphasia after stroke have been shown to retrieve words with high imageability more accurately than words with low imageability, as low imageability words are typically thought to be more difficult to process at the semantic level (e.g., Nickels & Howard, 1994; Luzzatti et al., 2002). However, opposite results have been found in the same population (Warrington, 1981), as well as in people with neurodegenerative diseases (Breedin, Saffran, & Coslett, 1994).

Motivation for the present study

There is little knowledge on the association of imageability ratings between and within languages. Imageability is a linguistic variable related to meaning. That is, it reflects the richness of the semantic representation of words (Breedin et al., 1994; Plaut & Shallice, 1993). Therefore, finding cross-linguistic correlations in imageability ratings between words that are semantically equivalent may indicate lexical/semantic similarities across languages. By semantic equivalence we mean words for which a language expert and proficient speaker of both languages provides a direct translation.

Concepts such as "apple" and "house" may be thought of as easy to imagine among speakers of the same language, but also among speakers of different languages. This is because they can be represented with a semantically equivalent word (e.g., "apple" and "mela" in Italian, or "house" and "kuća" in Serbian). At the same time, it could be argued that concepts that are dependent on cultural or socio-economic factors, such as "golf", "handrail" or "priest", may not have the same imageability ratings across languages. Along these lines, Blomberg and Öberg (2015) reported a strong positive correlation between English and Swedish imageability ratings. The authors argued that imageability ratings "can be reliably transferred between the two languages, although some caution should be taken, since for some individual words, some ratings might differ substantially" (p. 351).

If a positive finding for cross-linguistic similarities holds, existing imageability ratings in a widely studied language such as English may be used to norm and validate newly obtained ratings in a less studied language, or be used as approximate measures for the new language of interest. This could be useful at a practical level, as many languages possess few or no databases available that yield information on imageability (or other variables, for example, see Proctor & Vu, 1999). This lack of available ratings contrasts with a growing interest in empirical language studies and the need to adapt assessment materials to new languages (Fyndanis, Lind, Varlokosta, Simonsen, Kambanaros, Ceder et al., 2017). Therefore, such a finding could be used as an argument to utilize existing ratings of other languages and to speed up the adaptation of test materials into less researched languages. To the best of our knowledge, specific criteria to decide whether imageability ratings can be used across languages are non-existent. In this exploratory study, we assessed different criteria, including the number of semantically equivalent words between databases, the correlation value (rho), and the variance explained (R^2). Additionally, we discussed linguistic and cultural factors, intrinsic differences between databases, range effects, and mean age of participants.

In the present study, members the Collaboration of Aphasia Trialists (CATs) [COST IS1208] compared ratings of thirteen European languages (i.e., Basque, Catalan, Croatian, English, Greek, Cypriot Greek, Hungarian, Italian, Norwegian, Serbian, Spanish, Swedish, and Turkish). These data were collected as part of a project where we adapt The Comprehensive Aphasia Test into a range of languages spoken in Europe (Fyndanis et al., 2017). We expected to find strong positive correlations between the imageability ratings collected for different languages, provided that the words entered in the correlations are semantically equivalent.

Methods

Fifteen imageability databases were considered. These corresponded to unpublished data for 10 different languages, namely, Basque, Catalan, Croatian, Greek, Cypriot Greek, Hungarian, Serbian, Spanish, Swedish, and Turkish. We also included four published imageability ratings: three English datasets (Bird et al., 2001; Coltheart, 1981; Cortese & Fugett, 2004), one Italian (Rofes et al., 2015) and one Norwegian (Lind et al., 2015; Simonsen et al., 2013). Detailed information on the total number of words, informant characteristics, modality, scale used, and references for the published databases can be found in Table 1.

[Please add Table 1 around here]

Some differences existed between databases. The number of participants ranged between 20 and 399 and the mean age of participants between 21 and 65 years. Eight of 15 databases were collected using a web-based survey and six with a paper-based survey. The Greek database was collected with a paper-based and a web-based survey. Thirteen of the databases were collected using a 7-point scale and two databases using a 5-point scale. Furthermore, the Hungarian database only included nouns. Also, the Norwegian database included nouns, verbs and adjectives, but only the nouns were used in this study.

Instructions

Imageability ratings in all languages were obtained following the instructions by Paivio et al. (1968): The purpose of this experiment is to rate a list of words as to the ease or difficulty with which they arouse mental images. Any word which, in your estimation, arouses a mental image (i.e., a mental picture, or sound, or other sensory experience) very quickly and easily should be given a high imagery [imageability] rating; any word that arouses a mental image with difficulty or not at all should be given a low imagery [imageability] rating. Think of the worlds "apple" or "fact". Apple would probably arouse an image relatively easily and would be rated as high [imageability]; fact would probably do so with difficulty and would be rated as low [imageability]. (p.4)

Semantically equivalent words

The number of semantically equivalent words between each of the two English databases and each of the other languages ranged between four and 467. Semantic equivalence between two words was determined as follows: the relevant words for each language were listed, and for each word a language expert (native speaker of the language) indicated a corresponding English word equivalent. For example, for the word "poma" in Catalan, the English equivalent "apple" was given.

Statistical analyses

We correlated (i.e., Spearman rho coefficient) semantically equivalent words between all databases based on their English translation. We excluded all correlations where there were less than 20 words in common, as correlations with very low data points are vulnerable to error (e.g., Bonnet & Wright, 2000). We calculated 105 correlations and excluded 36 because they contained less than 20 words in common. In total, we included 69 correlations. The correlation across the English databases of Bird et al. (2001) and Cortese and Fugett (2004) had the greatest number of semantically equivalent words. This was followed by the English databases of Cortese and Fugett (2004) and Coltheart (1981), and the English database of Cortese and Fugett (2004) with the Norwegian database (467, 296, 251 semantically equivalent words, respectively). We also measured the amount of variation that could be explained by the relationship between each pair of databases. This is called variance explained (R^2). For example, given that Basque and Catalan correlate at a rho =.74, the variance in Basque is "explained" or predicted by the Catalan database by 55%. We have calculated this variance using the following formula: $rho^2 \times 100 = \%variance$ (in the example $.74^2 \times 100 = 55\%$).

Results

A summary of the correlations for the lists of semantically equivalent words across languages can be found in Table 2. A full description of each of the correlations including mean imageability for each set of semantically equivalent words, number of equivalent words, rho, p-value, and variance explained (R^2) can be found in the appendices (Table S1).

[Please add Table 2 around here]

We obtained 65 significant correlations, and four did not lead to significant results (i.e., English [Bird et al.] and Turkish; English [Bird et al.] and Catalan; English [MRC database] and Spanish; Basque and Hungarian). The strength of the correlations ranged from low (rho =.31 for Norwegian and Turkish) to high (rho =.92 for Catalan and Turkish) and had a moderate median value (rho =.68). The variance explained ranged from 9% (for Norwegian and Turkish) to 85% (for Catalan and Turkish) and had a median value of 46%. A matrix scatter plot representing the variability in number of semantically equivalent words and R^2 across datasets can be found in Figure 1.

[Please add Figure 1 around here]

Discussion

Imageability ratings have been collected, studied, and used to control experimental conditions in numerous psycholinguistic, neuropsychological, and aphasiological studies (e.g., Hanley & Kay, 1997; Kambanaros et al., 2013; Nickels & Howard, 1994; Paivio et al., 1968; Wise et al., 2000). Excellent work has been put forward to understand the consistency of such ratings within and between languages (i.e., Bird et al., 2001; Blomberg & Öberg, 2015; Cortese & Fugett, 2004; Simonsen et al., 2013). Yet, and to the best of our knowledge, no studies had considered this issue across multiple languages and using the same instructions to collect imageability ratings (Paivio et al., 1968). In this study, members of the Collaboration of Aphasia Trialists (CATs) [COST IS1208] addressed this issue across 13 European languages. Imageability ratings often correlated across languages. The median strength across correlations was moderate (rho =.68) and the variance explained reached 46%. This implies that at least 54% of the variation in this dataset was due to factors other than imageability.

Finding significant correlations across databases can be explained by the fact that imageability is a linguistic variable that reflects the richness of the semantic representation of a word (Breedin et al., 1994; Plaut & Shallice, 1993), and such representation should be relatively similar within and between languages (e.g., Bird et al., 2001; Cortese & Fugett, 2004; Cortese et al., 2012; Blomberg & Öberg, 2015). Even though associations across languages possibly exist, our current results should be interpreted cautiously as the moderate strength of the correlations and the variance explained indicate that there is no clear one-to-one correspondence between the imageability ratings across languages. We discuss some possible explanations for this remaining unexplained variance, including linguistic and cultural factors; intrinsic differences between databases; range effects; number of words, equivalent words, and participants; and mean age of participants.

Linguistic and cultural factors

The fact that words are semantically equivalent does not necessarily imply that these words also share similar ratings for other psycholinguistic variables such as frequency of usage, age of acquisition, word length, regularity of spelling, etc. Blomberg and Öberg (2015) found that the English word "sorrow" is higher in imageability but lower in frequency than its Swedish semantic equivalent "sorg"; the English word "anger" is less imageable than the Swedish semantic equivalent "ilska"; and that the English word "position" is lower in age of acquisition than the Swedish semantic equivalent "position". Another factor, which possibly affects the results, albeit to a lesser extent given that we only compared ratings from European speakers, is that ratings for imageability (and other variables) may also depend on the cultural setting. Simonsen et al. (2013) pointed out that: Most Norwegian children have to swallow a spoonful of cod liver oil every day at least through the winter months. It is fair to assume that the Norwegian word for cod liver oil, tran, has a high imageability compared to languages spoken in countries where this is not the custom. (p. 436)

Matching words for a series of linguistic variables was not possible in the present study. This is because some of the languages did not have norms for all these linguistic variables and, when they did, some of the databases did not include ratings for all the words. Matching all words for these psycholinguistic variables would have minimized the number of words entered in the study, hence, also reducing the overall power of the statistical analyses. Having said that, we performed a secondary analysis on age of acquisition ratings for languages that in this study contained ≥ 100 semantically equivalent words for imageability. The analyses comprised Basque (Duñabeitia, Casaponsa, Dimitropoulou, Martí, Larraza, & Carreiras, in preparation); two English databases (Bird et al., 2001; Cortese & Khanna, 2008; Schock, Cortese & Khanna, 2012); Norwegian (Simonsen et al., 2013; Lind et al. 2015); Italian (Rofes et al., 2015) and Spanish (Alonso, Fernandez, & Díez, 2015). The results indicated that, also for age of acquisition, the median strength across correlations was moderate (rho =.53) and the variance explained reached 28%. This implies that at least 72% of the variation in these datasets was due to factors other than age of acquisition. Further details can be found in the supplementary materials (Table S2).

Intrinsic differences between databases

Despite the fact that all the imageability ratings in all the databases were collected following the instructions by Paivio et al. (1968), many of the databases did not take into account rating differences that may appear due to the fact that participants may not know the grammatical category of the words they were rating. For example, if a participant is presented with the English word "brief" she may not know whether it is a noun or a verb, unless it is read as "a brief" or "to brief" – in fact, "brief" could also be an adjective as in "a brief history". The same holds for Norwegian, where a participant having to rate the word "føde", may either consider it a verb "give birth" or a noun "food", unless the infinitive marker "å" is used, as in "å føde" (to give birth).

Disambiguating cases of homonymy is relevant because nouns and verbs have different imageability and other psycholinguistic values (e.g., Howard & Franklin, 1988; Whitworth et al., 2014). This aspect is particularly relevant for some languages such as English and Norwegian. And, in fact, it was taken into account in the English database of Bird et al. (2001) and also in the Norwegian database (Lind et al., 2015; Simonsen et al., 2013). This specific aspect may not be as relevant for other languages, as the difference between nouns and verbs is marked morphologically and orthographically. Catalan and Spanish infinitives, for example, are marked with –(a/e/i)r, as in "cantar" (to sing). Also, Turkish infinitives are marked with –m(e/a)k, as in "bakmak" (to look) and "almak" (to take). A priori, having controlled for this factor makes the English database of Bird et al. (2001) and the Norwegian database different from the other databases. In this study, we noticed no special patterns regarding homonymy. English correlated with almost all databases except for Greek and Serbian which are languages in which homonymy is not an issue, as very few noun and verb homonyms exist. Additionally, Greek and Serbian correlated with Basque and with Norwegian. Basque is also a language with very few noun and verb homonyms, and the Norwegian database was controlled for homonymy (Lind et al., 2015; Simonsen et al., 2013).

Range effects

When finding semantically equivalent words between languages, the number of words that was entered in the correlation may have clustered around specific parts of the distribution (e.g., Poulton, 1975). In our study, this implies correlating subsets where the majority of words had been rated as high in imageability. This is a reasonable explanation for some of the non-significant correlations, as many of the databases were collected as part of another project that aims at adapting a language battery that includes highly picturable items (e.g., for object naming tasks) into multiple European languages (Fyndanis et al., 2017). If this was the only contributing factor, however, it would be hard to explain why a database such as that of Croatian, as opposed to other databases that were collected as part of that project, would significantly correlate with all databases. Indeed, to avoid range effects for those databases that were collected anew, we instructed each language team to include 20 to 100 items that were expected to produce low imageability ratings, as based on items in the database of Bird et al. (i.e., 2 to 3 points of 7 in imageability). Also, those databases that were already collected contained larger number of words and, therefore, included a wider range of imageability scores. Finally, range effects may not be accounted by the fact that the Croatian and Spanish database used a 5-point scale, as opposed to the rest of databases that used a 7-point scale. This is because the two scales produce similar results, for example, they share the same mean score when re-scaled (Dawes, 2012).

Number of words, equivalent words, and participants

The relatively small number of words that some of the databases contained (cf. 9240 words in the MRC Psycholinguistic database vs. 202 words in the Catalan database) diminished the potential number of equivalent words between databases. We minimized the effects of this factor by only considering those correlations between languages where we found at least 20 equivalent words in common. This resulted in an exclusion of 36 out of 105 correlations. Also, there is potentially unexplained variability in the imageability ratings for each word due to the varying numbers of participants in each survey. Again, in those databases that were collected anew, we tried to minimize this factor by including at least 20 individuals in each survey (e.g., Basque = 43 participants; Greek = 118 participants; Spanish = 20 participants; Swedish = 52 participants; Turkish = 51 participants).

Mean age of participants

The mean age of participants between databases is higher in the study of Bird et al. (2001) than in many of the other studies. If we take the mean age of participants as a factor, we see that the Catalan and the Turkish databases were rated by people around 20-25 years of age (be it undergraduate students or not), while the Basque, Hungarian, Greek and Swedish databases were rated by populations with a mean value of 40-45 years of age. This latter value could be thought of as closer to the 65 years of age of the database by Bird et al. (2001). This could explain some of the differences in significance testing, for example, the fact that the English database of Bird et al. (2001) did not correlate with the Catalan and Turkish databases. These results would be in line with an effect of age found in the Norwegian imageability study - from age 30 and upwards the imageability ratings increased systematically and significantly with participant age, with the largest difference found between 40 and 50 years (Simonsen et al, 2013). In the same vein, Bird et al. (2001) indicated that specific word ratings for some variables, such as age of acquisition, may differ depending on the age of the participants.

Additionally, it could be argued that older individuals may have richer semantic representations, due to experience, as vocabulary scores increase with age (Diaz et al., 2014). If this holds, Catalan would have obtained lower imageability scores Hungarian, Greek, and Swedish. However, on average, Hungarian, Greek, and Swedish were approximately 2 points lower in imageability than Catalan, this being against the fact that older participants provide higher imageability scores than younger participants. Additionally, Turkish scores were very similar to Hungarian, Greek, and Swedish (range=5.55-6.43, see supplementary materials). This is also against the older individuals providing higher imageability scores than younger individuals. Despite these results, it is worth noting that none of our databases included ratings from people of age 70 or older. Obtaining the ratings of people of age 70 or older may be relevant, as a decline in vocabulary scores has been reported in these individuals (Alwin & McCammon, 2001), and such a decline may be related to differences in imageability.

Future directions

A future study may consider a smaller number of languages and words matched for a series of variables (frequency, age of acquisition, linguistic typology, cultural factors) using the same methods. The study could assess whether or not the strength and variance explained in the correlations is higher when these variables are considered for word selection as opposed to when these variables are not considered for word selection. To the best possible extent, such a study may also avoid including words that are obviously strongly dependent on cultural factors. It could also be interesting to study how bilingual and multilingual speakers conceptualize the imageability of specific words and also to look at speakers with different levels of literacy/education. Given our current results, we would expect that these speakers would rate words with the same meaning indistinctively, regardless of the language, although some differences could emerge relative to literacy/education.

Conclusion

The high number of significant correlations between databases indicates that imageability ratings are, to a large extent, similar across languages. We argue in favor of similarities in imageability between databases and discuss different reasons for the moderate strength between the correlations and the low variance explained. All these reasons possibly interact in our dataset. In sum: these are exciting results from a practical perspective, as they suggest that imageability ratings from one language may be used in another language. However, more accurate results may be obtained when collecting scores for each individual language.

References

Alario, F. X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers, 36*(1), 140-155.

Allport, D. A., & Funnell, E. (1981). Components of the mental lexicon. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 295*(1077), 397-410.

Alonso, M. A., Fernández, A., & Díez, E. (2015). Subjective age-of-acquisition norms for 7,039 Spanish words. *Behavior Research Methods, 47*(1), 268-274.

Alwin, D. F., & McCammon, R. J. (2001). Aging, cohorts, and verbal ability. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 56*(3), S151-S161.

Berndt, R. S., Haendiges, A. N., Burton, M. W., & Mitchum, C. C. (2002). Grammatical class and imageability in aphasic word production: Their effects are independent. *Journal of Neurolinguistics, 15*(3), 353-371.

Bedny, M., & Thompson-Schill, S. L. (2006). Neuroanatomically separable effects of imageability and grammatical class during single-word comprehension. *Brain and Language, 98*(2), 127-139.

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*(1), 73-79.

Bleasdale, F. A. (1987). Concreteness-dependent associative priming: Separate lexical organization for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(4), 582-594.

Blomberg, F., & Öberg, C. (2015). Swedish and English word ratings of imageability, familiarity and age of acquisition are highly correlated. *Nordic Journal of Linguistics, 38*(03), 351-364.

Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Psychometrika, 65*(1), 23-28.

Breedin, S. D., Saffran, E. M., & Coslett, H. B. (1994). Reversal of the concreteness effect in a patient with semantic dementia. *Cognitive Neuropsychology, 11*(6), 617-660.

Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *The Quarterly Journal of Experimental Psychology, 25*(1), 85-95.

Chiarello, C., Shears, C., & Lund, K. (1999). Imageability and distributional typicality measures of nouns and verbs in contemporary English. *Behavior Research Methods, Instruments, & Computers, 31*(4), 603-637.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology, 33*(4), 497-505.

Coltheart, V., Laxon, V. J., & Keating, C. (1988). Effects of word imageability and age of acquisition on children's reading. *British Journal of Psychology, 79*(1), 1-12.

Cortese, M. J., Khanna, M. M., & Hacker, S. (2010). Recognition memory for 2,578 monosyllabic words. *Memory, 18*(6), 595-609.

Cortese, M. J., McCarty, D. P., & Schock, J. (2015). A mega recognition memory study of 2897 disyllabic words. *The Quarterly Journal of Experimental Psychology, 68*(8), 1489-1501.

Cortese, M. J., & Schock, J. (2013). Imageability and age of acquisition effects in disyllabic word recognition. *The Quarterly Journal of Experimental Psychology, 66*(5), 946-972.

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers, 36*(3), 384-387.

Davelaar, E., & Besner, D. (1988). Word identification: Imageability, semantics, and the content-functor distinction. *The Quarterly Journal of Experimental Psychology, 40*(4), 789-799.

Dawes, J. G. (2012). Do data characteristics change according to the number of scale points used? An experiment using 5 point, 7 point and 10 point scales. *International Journal of Market Research, 50*(1), 61-77.

de Aguiar, V., Bastiaanse, R., Capasso, R., Gandolfi, M., Smania, N., Rossi, G., & Miceli, G. (2015). Can tDCS enhance item-specific effects and generalization after linguistically motivated aphasia therapy for verbs? *Frontiers in Behavioral Neuroscience, 9*, 190.

Della Rosa, P., 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, 42*, 1042–1048

Desrochers, A., & Thompson, G. L. (2009). Subjective frequency and imageability ratings for 3,600 French nouns. *Behavior Research Methods, 41*, 546–557.

Diaz, M. T., Johnson, M. A., Burke, D. M., & Madden, D. J. (2014). Age-related differences in the neural bases of phonological and semantic processes. *Journal of Cognitive Neuroscience, 26*(12), 2798-2811.

Duñabeitia, J.A., Casaponsa, A., Dimitropoulou, M., Martí, A., Larraza, S., & Carreiras, M. (In preparation). BaSp: A Basque-Spanish database of translation equivalents.

Franklin, S., Howard, D., & Patterson, K. (1995). Abstract word anomia. *Cognitive Neuropsychology, 12*(5), 549-566.

Fyndanis, V., Lind, M., Varlokosta, S., Simonsen, H.G., Kambanaros, M., Ceder, K., Rofes, A., Soroli, E., Bjekić, J., Gavarró, A., Grohmann, K., Kuvač Kraljević, J., Martínez-Ferreiro, S., Munarriz, A., Pourquie, M., Vuksanović, J., Zakariás, L., & Howard, D. (2017). Cross-linguistic adaptations of *The Comprehensive Aphasia Test*: Challenges and solutions. Clinical Linguistics and Phonetics, 1-14.

Hanley, R. J., & Kay, J. (1997). An effect of imageability on the production of phonological errors in auditory repetition. *Cognitive Neuropsychology, 14*(8), 1065-1084.

Howard, D., & Franklin, S. (1988). *Missing the meaning? A cognitive neuropsychological study of the processing of words by an aphasic patient.* London: MIT Press.

Kambanaros, M., & Grohmann, K. K. (2015). Grammatical class effects across impaired child and adult populations. *Frontiers in Psychology, 6*, 1-17.

Kambanaros, M., Grohmann, K. K., & Michaelides, M. (2013). Lexical retrieval for nouns and verbs in typically developing bilectal children. *First Language, 33*(2), 182-199.

Kuvac Kraljevic, J. & Olujic, M. (subm) Croatian Lexical Database.

Law, S. P., Kong, A. P. H., Lai, L. W. S., & Lai, C. (2015). Effects of context and word class on lexical retrieval in Chinese speakers with anomic aphasia. *Aphasiology, 29*(1), 81-100.

Levelt, W. (2014). *A history of psycholinguistics: The pre-Chomskyan era*. Oxford University Press.

Lind, M., Simonsen, H.G., Hansen, P., Holm, E. & Mevik, B.-H. 2015. Norwegian Words: A lexical database for clinicians and researchers. *Clinical Linguistics & Phonetics, 29*, 276-290.

Luzzatti, C., Raggi, R., Zonca, G., Pistarini, C., Contardi, A., & Pinna, G. D. (2002). Verb–noun double dissociation in aphasic lexical impairments: The role of word frequency and imageability. *Brain and Language, 81*(1), 432-444.

Ma, W., Golinkoff, R. M., Hirsh-Pasek, K., McDonough, C., & Tardif, T. (2009). Imageability predicts the age of acquisition of verbs in Chinese children. *Journal of Child Language, 36*, 405-423

McMullen, P. A., & Bryden, M. P. (1987). The effects of word imageability and frequency on hemispheric asymmetry in lexical decisions. *Brain and Language, 31*(1), 11-25.

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(1), 116-133.

Nickels, L., & Howard, D. (1994). A frequent occurrence? Factors affecting the production of semantic errors in aphasic naming. *Cognitive Neuropsychology, 11*(3), 289-320.

Paivio, A. (2014). Intelligence, dual coding theory, and the brain. *Intelligence, 47*, 141-158.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76*(1), 1-25.

Patterson, K. E. (1981). Neuropsychological approaches to the study of reading. *British Journal of Psychology, 72*(2), 151-174.

Plaut, D. C., & Shallice, T. (1993). Preservative and semantic influences on visual object naming errors in optic aphasia: A connectionist account. *Journal of Cognitive Neuroscience, 5*(1), 89-117.

Poulton, E.C. (1975). Range effects in experiments on people. *The American Journal of Psychology, 88*(1), 3-32.

Proctor, R. W., & Vu, K. P. L. (1999). Index of norms and ratings published in the Psychonomic Society journals. *Behavior Research Methods, Instruments, & Computers, 31*(4), 659-667.

Rofes, A., Capasso, R., & Miceli, G. (2015). Verb production tasks in the measurement of communicative abilities in aphasia. *Journal of Clinical and Experimental Neuropsychology, 37*(5), 483-502.

Rofes, A., de Aguiar, V., & Miceli, G. (2015). A minimal standardization setting for language mapping tests: An Italian example. *Neurological Sciences, 36*(7), 1113-1119.

Nishimoto, T., Ueda, T., Miyawaki, K., Une, Y., & Takahashi, M. (2012). The role of imagery-related properties in picture naming: A newly standardized set of 360 pictures for Japanese. *Behavior Research Methods, 44*, 934-945.

Sabsevitz, D. S., Medler, D. A., Seidenberg, M., & Binder, J. R. (2005). Modulation of the semantic system by word imageability. *Neuroimage, 27*(1), 188-200.

Schock, J., Cortese, M. J., & Khanna, M. M. (2012). Imageability estimates for 3,000 disyllabic words. *Behavior Research Methods, 44*(2), 374-379.

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language, 27*(5), 499-520.

Simonsen, H. G., Lind, M., Hansen, P., Holm, E., & Mevik, B. H. (2013). Imageability of Norwegian nouns, verbs and adjectives in a cross-linguistic perspective. *Clinical Linguistics & Phonetics, 27*(6-7), 435-446.

Stoke, S. M. (1929). Memory for onomatopes. *The Pedagogical Seminary and Journal of Genetic Psychology, 36*(4), 594-596.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(5), 1140.

Tyler, L. K., & Moss, H. E. (1997). Imageability and category-specificity. *Cognitive Neuropsychology, 14*(2), 293-318.

Tyler, L. K., Moss, H. E., Galpin, A., & Voice, J. K. (2002). Activating meaning in time: The role of imageability and form-class. *Language and Cognitive Processes, 17*(5), 471-50

Warrington, E. K. (1981). Concrete word dyslexia. *British Journal of Psychology, 72*(2), 175-196.

Whitworth, A., Webster, J., & Howard, D. (2014). *A cognitive neuropsychological approach to assessment and intervention in aphasia: A clinician's guide*. Psychology Press.

Table 1 Number of words, participant characteristics, and modality to obtain the data per language

Language	Total words	Participants	Modality	Scale	Reference
Basque	260	43 (mean age=42, SD=17)	Web-based	7-point	
Catalan	202	32 (university undergraduates)	Web-based	7-point	
Croatian	608	27-46 (mean age=44, SD=18)	Web-based	5-point	Kuvač Kraljević & Olujić, subm.
English	2020	78 (mean age=65; SD=9)	Paper-based	7-point	Bird et al., 2001
English	9240	Various databases, not reported	Paper-based	7-point	Coltheart, 1981
English	3000	31 (university undergraduates)	Paper-based	7-point	Cortese & Fugett, 2004
Greek	76	118 (mean age=42, SD=10.2)	Paper & web-based	7-point	
Cypriot Greek	80	40 (mean age=39; SD=14)	Paper-based	7-point	
Hungarian	207	31-37 (mean age=44, SD=12)	Web-based	7-point	
Italian	292	50 (mean age=28, SD=11)	Web-based	7-point	Rofes et al., 2015
Norwegian	917	399 (mean age=38, SD=16)	Web-based	7-point	Lind et al., 2015; Simonsen et al., 2013
Serbian	82	30 (mean age=31, SD=12)	Paper-based	7-point	
Spanish	256	20 (mean age=22, SD=5)	Web-based	5-point	
Swedish	190	52 (mean age=41, SD=17)	Web-based	7-point	
Turkish	176	22-29 (mean age=21; SD=1)	Paper-based	7-point	

Notes. In the Croatian database the values 27-46 indicate the range of participants that rated each word, as different words were rated by different numbers of participants. In the Norwegian database, not all participants rated all the words. That is, "the mean number of ratings for each word in the database [was] 23.5, with a standard deviation of 2.7. The range of ratings [was] 11–52. (Simonsen et al., 2013: 439)." In the Hungarian database, 31-37, and in the Turkish database, 22-29, indicate the number of participants that rated each list, as two lists were used.

Table 2 Number of semantically equivalent words and variance explained (R2) for correlations

	Basque	Catalan	Croatian	Eng_ Bird	Eng_ Cortese	Eng_ MRC1	Greek	Cypriot_ Greek	Hungarian	Italian	Norwegian	Serbian	Spanish	Swedish	Turkish
Basque															
Catalan	35*; 55%														
Croatian	110*; 42%	46*; 56%													
Eng_ Bird	56*; 38%	20; 15%	100*; 49%												
Eng_ Cortese	152*; 48%	64*; 37%	162*; 67%	467*; 46%											
Eng_ MRC1	61*; 31%	23*; 77%	91*; 62%	105*; 47%	296*; 76%										
Greek	26*; 28%	6; NA	23*; 61%	12; NA	18; NA	17; NA									
Cypriot_ Greek	20*; 43%	6; NA	25*; 67%	23*; 30%	22*; 50%	13; NA	64*; 76%								
Hungarian	26; 8%	11; NA	35*; 45%	24; NA	73*; 9%	28*; 49%	10; NA	11; NA							
Italian	44*; 16%	9; NA	46*; 37%	111*; 74%	139*; 45%	33*; 76%	6; NA	7; NA	20*; 69%						
Norwegian	145*; 24%	44*; 27%	145*; 35%	133*; 38%	251*; 42%	135*; 25%	27*; 38%	31*; 28%	66*; 66%	70*; 19%					
Serbian	28*; 20%	5; NA	61*; 64%	15; NA	31; 25%	15; NA	7; NA	7; NA	4; NA	11; NA	34*; 40%				
Spanish	37*; 58%	10; NA	29*; 52%	45*; 35%	102*; 32%	27; 12%	4; NA	4; NA	14; NA	13; NA	39*; 35%	7; NA			
Swedish	31*; 13%	18; NA	43*; 47%	27*; 53%	75*; 55%	28*; 48%	12; NA	12; NA	12; NA	19; NA	64*; 60%	12; NA	9; NA		
Turkish	66*; 28%	28*; 85%	83*; 47%	31; 9%	65*; 38%	45*; 43%	13; NA	15 NA	19; NA	30*; 41%	84*; 9%	19; NA	20*; 67%	24*; 48%	

Eng_Bird = data from Bird et al. (2001); Eng_Cortese = data from Cortese and Fugett (2004); Eng_MRCI1 = English data from Coltheart (1981). We only reported results for correlations with ≥ 20 semantically equivalent words. Significant correlations are marked with an asterisk (*).

Figure 1 Matrix scatterplot representing the variability across datasets

From top left to bottom right: English (Cortese) and English (MRC) as an example of a correlation between the same language with a large number of semantically equivalent words and high variance explained; English (Bird) and Norwegian as an example with a large number of semantically equivalent words, high variance explained, and data collected by asking participants to explicitly differentiate word category (nouns v verbs); English (Bird) and Spanish as an example with average number of semantically equivalent words, moderate variance explained and data collection for English by explicitly differentiating word category (nouns v verbs), and Spanish, as a language differentiates word category in the word form; Norwegian and Turkish for an average number of equivalent words and low variance explained; Hungarian and Italian for a low number of equivalent words and a relatively high variance explained; and Basque and Serbian for a low number of equivalent words and low variance explained.

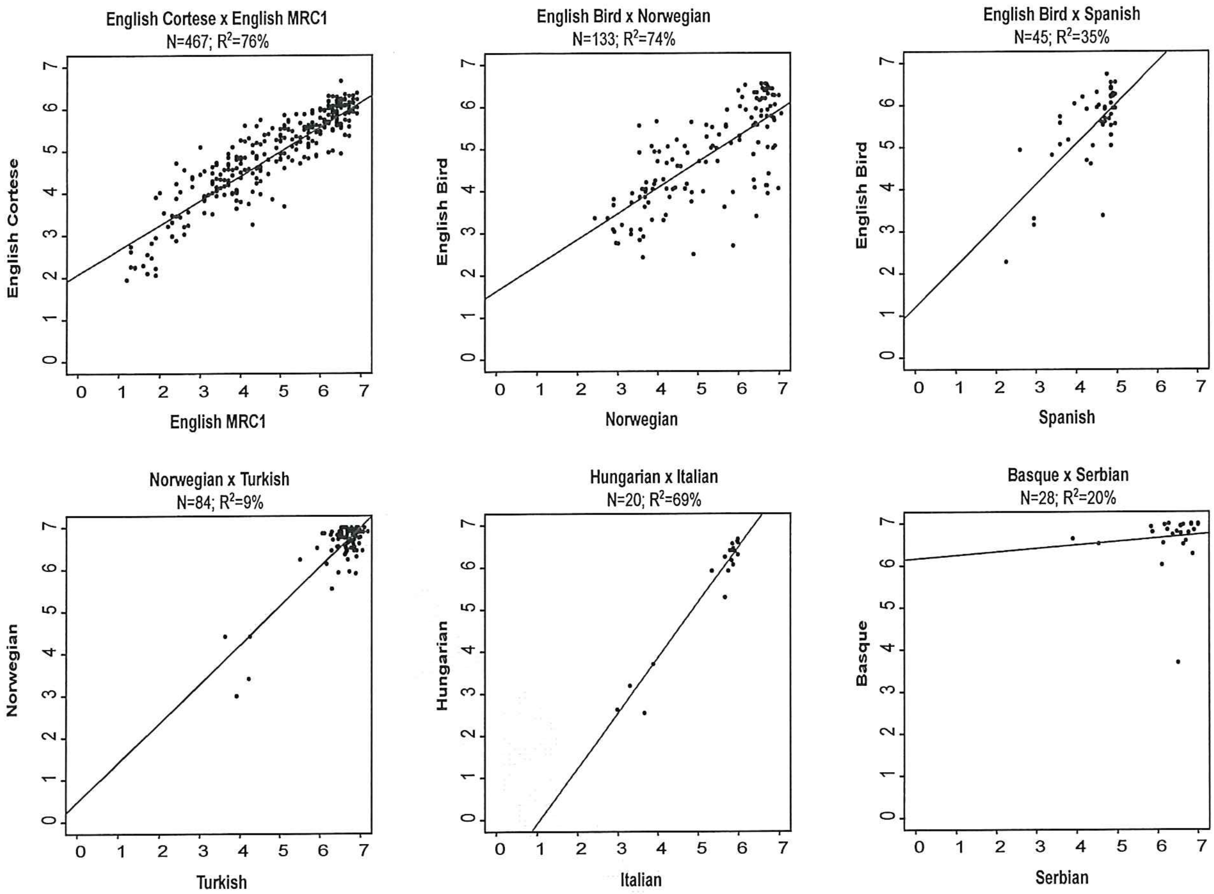
English Cortese x English MRC1
N=467; R²=76%
English Cortese
English MRC1
English Bird x Norwegian
N=133; R²=74%
English Bird
Norwegian
English Bird x Spanish
N=45; R²=35%
English Bird
Spanish
Norwegian x Turkish
N=84; R²=9%
Norwegian
Turkish
Hungarian x Italian
N=20; R²=69%
Hungarian
Italian
Basque x Serbian
N=28; R²=20%
Basque
Serbian

Table S1. Mean imageability, number of equivalent words between languages, rho, p-values and variance explained (R^2) for correlations

	Basque	Catalan	Croatian	Eng_Bird	Eng_ Cortese	Eng_ MRC1	Greek	Cypriot_ Greek	Hungarian	Italian	Norwegian	Serbian	Spanish	Swedish	Turkish
Basque	-	Basque: 6.2 (1.43) Catalan: 1.67 (1.16) 35; -.74; 3.617e- 07* 55%	Basque: 6.29 (1.29) Croatian: 4.62 (0.65) 110; .65; 1.161e- 14* 42%	Basque: 5.68; (1.59) Eng_Bird: 5.18 (1.09) 56; .62; 4.038e- 07* 38%	Basque: 6.33 (1.1) Eng_Corte se: 5.96 (1.07) 152; .69; 2.2e16* 48%	Basque: 6.48 (1.2) English_M RC1: 5.58 (123) 61; .56; 2.232e- 06* 31%	Basque: 6.32(1.49) Greek: 6.49 (0.86) 26; .53; 0.03* 28%	Basque: 6.39 (1.42) Cypriot_G reek: 6.67 (0.51) 20; .65; 0.001* 43%	Basque: 6.31 (1.31) Hungarian : 5.98 (1.16) 26; .29; 0.14 8%	Basque: 6.76 (0.38) Italian: 5.79 (0.38) 44; .4; 0.01* 16%	Basque: 6.57 (0.84) Norwegian : 6.48 (0.67) 145; .49; 4.526e- 10* 24%	Basque: 6.68 (0.63) Serbian: 6.38 (0.7) 28; .45; 0.02* 20%	Basque: 6.79 (0.25) Spanish: 4.64 (0.3) 37; .76; 6.505e- 08* 58%	Basque: 6.48 (1.27) Swedish: 6.57 (0.94) 31; .35; 0.05* 13%	Basque: 6.42 (1.32) Turkish;: 6.43 (0.86) 66; .53; 5.398e- 06* 28%
Catalan	Catalan: 1.67 (1.16) Basque: 6.2 (1.43) 35; -.74; 3.617e-07* 55%	-	Catalan: 4.1 (1.12) Croatian: 2.44 (1.8) 46; -.79; 4.895- e.11* 56%	Catalan: 5.17 (1.77) Eng_Bird: 4.48 (1.19) 20; .38; 0.09 15%	Catalan: 5.28 (1.75) Eng_Corte se: 5.23 (1.78) 64; .61; 5.828e- 08* 37%	Catalan: 2.99 (2.06) English_M RC1: 4.74(1.58) 23; -.88; 2.954.e- 08* 77%	Catalan: 5.82 (1.41) Greek: 2.26 (1.76) 6; NA	Catalan: 2.98 (2.15) Cypriot_G reek: 5.94 (1.07) 6; NA	Catalan: 5.12 (1.76) Hungarian : 2.47 (1.74) 11; NA	Catalan: 5.28 (1.02) Italian: 2.28 (1.92) 9; NA	Catalan: 6.38 (0.93) Norwegian : 1.61 (1.23) 44; -.52; 0.002* 27%	Catalan: 6.15 (0.54) Serbian: 1.09 (0.02) 5; NA	Catalan: 4.39 (0.88) Spanish: 1.61 (1.21) 10; NA	Catalan: 5.98 (1.38) Swedish: 2.23 (1.79) 18; NA	Catalan: 5.64 (1.51) Turkish: 2.38 (1.67) 28; -.92; 1.65e-11* 85%
Croatian	Croatian: 4.62 (0.65) Basque: 6.29 (1.29) 110; .65; 1.161e-14* 42%	Croatian: 2.44 (1.8) Catalan: 4.1 (1.12) 46; -.79; 4.895- e.11* 56%	-	Croatian: 3.96 (0.98) Eng_Bird 4.84 (1.19) 100; .70; 2.242e- 14* 49%	Croatian: 4.29 (0.89) Eng_Corte se: 5.43 (1.44) 162; .82; 2.2-e-16* 67%	Croatian: 4.28 (0.86) English_M RC1: 5.13 (1.33) 91; .79; 2.2e-16* 62%	Croatian: 6-25 (1.08) Greek: 4.58 (0.66) 23; .78; 1.1865- 05* 61%	Croatian: 4.56 (0.76) Cypriot_G reek: 6.59 (0.62) 25; .82; 4.94e-07* 67%	Croatian: 5.56 (1.47) Hungarian : 4.37 (0.9) 35; .67; 1.22e-05* 45%	Croatian: 5.56 (0.69) Italian: 4.37 (0.83) 46; .61; 7.553e- 06* 37%	Croatian: 6.22 (0.99) Norwegian : 4.61 (0.65) 145; .59; 3.89e.15* 35%	Croatian: 5.87 (1.18) Serbian: 4.72 (0.33) 61; .80; 1.201e- 14* 64%	Croatian: 4.36 (0.83) Spanish: 4.51 (0.89) 29; .72; 1.165e- 05* 52%	Croatian: 4.63 (0.63) Swedish: 6.36 (0.95) 43; .69; 3.142e- 07* 48%	Croatian: 6.13 (1.18) Turkish: 4.42 (0.93) 83; .69; 2.171e- 13* 48%
Eng_Bird	Eng_Bird: 5.18 (1.09) Basque: 5.68; (1.59) 56; .62; 4.038e-07*	Eng_Bird: 4.48 (1.19) Catalan: 5.17 (1.77) 20; .38; 0.09	Eng_Bird 4.84 (1.19) Croatian: 3.96 (0.98) 100; .70; 2.242e-	-	Eng_Bird: 4.47 (1.01) Eng_Corte se: 4.22 (1.32) 467; .68;	Eng_Bird: 4.57 (1.13) English_M RC1: 4.53 (1.18) 105; .69;	Eng_Bird: 5.4 (0.76) Greek: 6.08 (0.98) 12; NA	Eng_Bird 5.1 (0.9) Cypriot_G reek: 6.21 (0.9) 23; .55;	Eng_Bird: 4.53 (1.29) Hungarian : 4.29 (1.47) 24; .71;	Eng_Bird: 4.43 (1.22) Italian: 4.74 (1.08) 111; .86; 2.2e-16*	Eng_Bird: 4.86 (1.14) Norwegian : 5.26 (1.35) 133; .86;	Eng_Bird: 5.49 (0.82) Serbian: 6.03 (1.02) 15 NA	Eng_Bird: 5.54 (0.95) Spanish: 4.39 (0.68) 45; .59; 2.26e-05*	Eng_Bird: 5.02 (1.09) Swedish: 5.65 (1.28) 27; .73; 1.224e-	Eng_Bird: 5.06 (1.12) Turkish: 6.23 (0.86) 31; .31 0.08

	38%	15%	14* 49%		2.2e-16* 46%	4.033e- 16* 47%		0.006* 30%	1.05e-04* 50%	74%	2.2e-16* 74%		35%	05* 53%	9%
Eng_Cortese	Eng_Cortese: 5.96 (1.07) Basque: 6.33 (1.1) 152; .69; 2.2e16* 48%	Eng_Corte se: 5.23 (1.78) Catalan: 5.28 (1.75) 64; .61; 5.828e- 08* 37%	Eng_Corte se: 5.43 (1.44) Croatian: 4.29 (0.89) 162; .82; 2.2-e-16* 67%	Eng_Corte se: 4.22 (1.32) Eng_Bird: 4.47 (1.01) 467; .68; 2.2e-16* 46%	-	Eng_Corte se 4.83(1.52) English_M RC1: 4.91 (1.12) 296; .87; 2.2e-16* 76%	Eng_Corte se: 5.67 (1.42) Greek: 6.41 (0.66) 18; NA	Eng_Corte se: 5.53 (1.55) Cypriot_G reek: 6.45 (0.66) 22; .71; 0.0001* 50%	Eng_Corte se: 4.94 (1.58) Hungarian : 4.81 (1.59) 73; .89; 5.53e-26* 79%	Eng_Corte se: 4.87 (1.28) Italian: 5.49 (0.57) 139; .67; 2.2e-16* 45%	Eng_Corte se: 5.64 (1.27) Norwegian : 5.92 (1.09) 251; .65; 2.2e-16* 42%	Eng_Corte se: 6.26 (0.54) Serbian: 6.39 (0.67) 31; .50; 0.003 25%	Eng_Corte se: 5.98 (1.15) Spanish: 4.52 (0.58) 102; .57; 3.219e- 10* 32%	Eng_Corte se: 5.43 (1.36) Swedish: 6.09 (1.02) 75; .74; 2.65e-14* 55%	Eng_Corte se: 5.85 (1.35) Turkish: 6.21 (1.09) 65; .62; 3.16e-08* 38%
Eng_MRC1	English_MR C1: 5.58 (123) Basque: 6.48 (1.2) 61; .56; 2.232e-06* 31%	English_M RC1: 4.74(1.58) Catalan: 2.99 (2.06) 23; -.88; 2.954.e- 08* 77%	English_M RC1: 5.13 (1.33) Croatian: 4.28 (0.86) 91; .79; 2.2e-16* 62%	English_M RC1: 4.53 (1.18) Eng_Bird: 4.57 (1.13) 105; .69; 4.033e- 16* 47%	English_M RC1: 4.91 (1.12) Eng_Corte se 4.83(1.52) 296; .87; 2.2e-16* 76%	-	English_M RC1: 5.54 (0.9) Greek: 5.97 (1.3) 17; NA	English_M RC1: 5.59 (0.91) Cypriot_G reek: 6.41 (0.77) 13; NA	English_M RC1: 4.88 (1.38) Hungarian : 4.98 (1.45) 28; .70; 2.912e- 05* 49%	English_M RC1: 4.93 (1.1) Italian: 5.26 (0.92) 33; .87; 3.236e- 11* 76%	English_M RC1: 5.37 (1.24) Norwegian : 6.11 (1.05) 135; .50; 9.775e- 10* 25%	English_M RC1: 5.9 (2.2) Serbian: 6.2 (0.73) 15; NA	English_M RC1: 5.73 (0.73) Spanish: 4.49 (0.56) 27; .35; 0.07 12%	English_M RC1: 5.18 (1.4) Swedish: 6.02 (1.1) 28; .69; 4.549e- 05* 48%	English_M RC1: 5.42 (1.23) Turkish: .25 (0.97) 45; .65; 1.59e-06* 43%
Greek	Greek: 6.49 (0.86) Basque: 6.32(1.49) 26; .53; 0.03* 28%	Greek: 2.26 (1.76) Catalan: 5.82 (1.41) 6; NA	Greek: 4.58 (0.66) Croatian: 6-25 (1.08) 23; .78; 1.1865- 05* 61%	Greek: 6.08 (0.98) Eng_Bird: 5.4 (0.76) 12; NA	Greek: 6.41 (0.66) Eng_Corte se: 5.67 (1.42) 18; NA	Greek: 5.97 (1.3) English_M RC1: 5.54 (0.9) 17; NA	-	Greek: 5.97 (1.09) Cypriot_G reek: 6.24 (0.85) 64; .87; 2.2e-16* 76%	Greek: 6.12 (0.96) Hungarian : 6.65 (0.41) 10; NA	Greek: 6.25 (1.22) Italian: 5.49 (0.76) 6; NA	Greek: 6.24 (1.14) Norwegian : 6.39 (0.97) 27; .62; 0.0005* 38%	Greek: 6.65 (0.38) Serbian: 6.86 (0.05) 7; NA	Greek: 4.85 (0) Spanish: 6.9 (0.6) 4; NA	Greek: 6.21 (1.24) Swedish 6.18 (1.23) 12; NA	Greek: 6.29 (0.87) Turkish: 6.34 (1.01) 13; NA
Cypriot_Greek	Cypriot_Gree k: 6.67 (0.51) Basque: 6.39 (1.42) 20; .65; 0.001* 43%	Cypriot_G reek: 5.94 (1.07) Catalan: 2.98 (2.15) 6; NA	Cypriot_G reek: 6.59 (0.62) Croatian: 4.56 (0.76) 25; .82; 4.94e-07* 67%	Greek_Cp yriot 6.21 (0.9) Eng_Bird 5.1 (0.9) 23; .55; 0.006* 30%	Cypriot_G reek: 6.45 (0.66) Eng_Corte se: 5.53 (1.55) 22; .71; 0.0001*	Cypriot_G reek: 6.41 (0.77) English_M RC1: 5.59 (0.91) 13; NA	Cypriot_G reek: 6.24 (0.85) Greek: 5.97 (1.09) 64; .87; 2.2e-16* 76%	-	Cypriot_G reek: 6.78 (0.21) Hungarian : 6.13 (0.91) 11; NA	Cypriot_G reek: 6.57 (0.69) Italian: 5.54 (0.71) 7; NA	Cypriot_G reek: 6.5 (0.75) Norwegian : 6.16 (1.17) 31; .53; 0.002*	Cypriot_G reek: 6.85 (0.13) Serbian: 6.66 (0.38) 7; NA	Cypriot_G reek: 6.87 (0.04) Spanish: 4.85 (0.1) 4; NA	Cypriot_G reek: 6.25 (0.93) Swedish: 6.65 (0.38) 12; NA	Cypriot_G reek: 6.57 (0.61) Turkish: 6.13 (1.16) 15; NA

					50%						28%				
Hungarian	Hungarian: 5.98 (1.16) Basque: 6.31 (1.31) 26; .29; 0.14 8%	Hungarian: 2.47 (1.74) Catalan: 5.12 (1.76) 11; NA	Hungarian: 4.37 (0.9) Croatian: 5.56 (1.47) 35; .67; 1.22e-05* 45%	Hungarian: 4.37 (0.9) Croatian: 5.56 (1.47) 24; NA	Hungarian: 4.81 (1.59) Eng_Corse: 4.94 (1.58) 73; .89; 5.53e-26* 9%	Hungarian: 4.98 (1.45) English_MRC1: 4.88 (1.38) 28; .70; 2.912e-05* 49%	Hungarian: 6.65 (0.41) Greek: 6.12 (0.96) 10; NA	Hungarian: 6.13 (0.91) Cypriot_Greek: 6.78 (0.21) 11; NA	-	Hungarian: 5.34 (0.99) Italian: 5.62 (1.39) 20; .83; 4.544e-06* 69%	Hungarian: 5.59 (1.54) Norwegian: 5.18 (1.65) 66; .81; 2.2e-16* 66%	Hungarian: 6.76 (0.3) Serbian 6.37 (0.06) 4; NA	Hungarian: 4.59 (0.72) Spanish: 6.13 (0.9) 14; NA	Hungarian: 5.45 (1.58) Swedish 4.93 (1.76) 12 NA	Hungarian: 6.04 (1.4) Turkish: 5.55 (1.43) 19; NA
Italian	Italian: 5.79 (0.38) Basque: 6.76 (0.38) 44; .40; 0.01* 16%	Italian: 2.28 (1.92) Catalan: 5.28 (1.02) 9; NA	Italian: 4.37 (0.83) Croatian: 5.56 (0.69) 46; .61; 7.553e-06* 37%	Italian: 4.74 (1.08) Eng_Bird: 4.43 (1.22) 111; .86; 2.2e-16* 74%	Italian: 5.49 (0.57) Eng_Corse: 4.87 (1.28) 139; .67; 2.2e-16* 45%	Italian: 5.26 (0.92) English_MRC1: 4.93 (1.1) 33; .87; 3.236e-11* 76%	Italian: 5.49 (0.76) Greek: 6.25 (1.22) 6; NA	Italian: 5.54 (0.71) Cypriot_Greek: 6.57 (0.69) 7; NA	Italian: 5.62 (1.39) Hungarian: 5.34 (0.99) 20; .83; 4.544e-06* 69%	-	Italian: 5.52 (0.77) Norwegian: 6.22 (1.15) 70; 0.44; 0.0001* 19%	Italian: 6.75 (0.42) Serbian 5.56 (0.71) 11; NA	Italian: 4.67 (0.63) Spanish: 5.73 (0.63) 13; NA	Italian: 5.81 (1.51) Swedish: 5.32 (1.04) 19; NA	Italian: 6.6 (0.6) Turkish: 5.78 (0.45) 30; .64; 9.799e-05* 41%
Norwegian	Norwegian: 6.48 (0.67) Basque: 6.57 (0.84) 145; .49; 4.526e-10* 24%	Norwegian: 1.61 (1.23) Catalan: 6.38 (0.93) 44; -.52; 0.002* 27%	Norwegian: 4.61 (0.65) Croatian: 6.22 (0.99) 145; .59; 3.89e.15* 35%	Norwegian: 5.26 (1.35) Eng_Bird: 4.86 (1.14) 133; .62; 3.16e-08* 38%	Norwegian: 5.92 (1.09) Eng_Corse: 5.64 (1.27) 251; .65; 2.2e-16* 42%	Norwegian: 6.11 (1.05) English_MRC1: 5.37 (1.24) 135; .51; 9.775e-10* 25%	Norwegian: 6.39 (0.97) Greek: 6.24 (1.14) 27; .62; 0.0005* 38%	Norwegian: 6.16 (1.17) Cypriot_Greek: 6.5 (0.75) 31; .53; 0.002* 28%	Norwegian: 5.18 (1.65) Hungarian: 5.59 (1.54) 66; .81; 2.2e-16* 66%	Norwegian: 6.22 (1.15) Italian: 5.52 (0.77) 70; .44; 0.0001* 19%	-	Norwegian: 6.45 (0.66) Serbian: 6.54 (0.47) 34; .63; 6.96e-05* 40%	Norwegian: 4.63 (0.44) Spanish: 6.52 (0.68) 39; .59; 7.735e-05* 35%	Norwegian: 5.79 (1.41) Swedish: 5.97 (1.35) 64; .78; 3.33e-14* 60%	Norwegian: 6.58 (0.69) Turkish: 6.54 (0.63) 84; .31 0.005* 9%
Serbian	Serbian: 6.38 (0.7) Basque: 6.68 (0.63) 28; .45; 0.02* 20%	Serbian: 1.09 (0.02) Catalan: 6.15 (0.54) 5; NA	Serbian: 4.72 (0.33) Croatian: 5.87 (1.18) 61; .80; 1.201e-14*	Serbian: 6.03 (1.02) Eng_Bird: 5.49 (0.82) 15; NA	Serbian: 6.39 (0.67) Eng_Corse: 6.26 (0.54) 31; .50; 0.003	Serbian: 6.2 (0.73) English_MRC1: 5.9 (2.2) 15; NA	Serbian: 6.86 (0.05) Greek: 6.65 (0.38) 7; NA	Serbian: 6.66 (0.38) Cypriot_Greek: 6.85 (0.13) 7; NA	Serbian 6.37 (0.06) Hungarian: 6.76 (0.3) 4; NA	Serbian 5.56 (0.71) Italian: 6.75 (0.42) 11; NA	Serbian: 6.54 (0.47) Norwegian: 6.45 (0.66) 34; .63; 6.96e-05*	-	Serbian: 4.66 (0.26) Spanish: 6.19 (0.72) 7; NA	Serbian: 6.66 (0.64) Swedish: 6.6 (0.32) 12; NA	Serbian: 6.64 (0.4) Turkish: 6.53 (0.44) 19; NA

	[Basque]	[Catalan]	[Croatian]	[Eng_Bird]	[Eng_Cortese]	[English_MRC1]	[Greek]	[Cypriot_Greek]	[Hungarian]	[Italian]	[Norwegian]	[Serbian]	[Spanish]	[Swedish]	[Turkish]
			64%		25%						40%				
Spanish	Spanish: 4.64 (0.3) Basque: 6.79 (0.25)	Spanish: 1.61 (1.21) Catalan: 4.39 (0.88)	Spanish: 4.51 (0.89) Croatian: 4.36 (0.83)	Spanish: 4.39 (0.68) Eng_Bird: 5.54 (0.95)	Spanish: 4.52 (0.58) Eng_Cortese: 5.98 (1.15)	Spanish: 4.49 (0.56) English_MRC1: 5.73 (0.73)	Spanish: 6.9 (0.6) Greek: 4.85 (0)	Cypriot_Greek: 6.87 (0.04) Spanish: 4.85 (0.1)	Spanish: 6.13 (0.9) Hungarian: 4.59 (0.72)	Spanish: 5.73 (0.63) Italian: 4.67 (0.63)	Spanish: 6.52 (0.68) Norwegian: 4.63 (0.44)	Spanish: 6.19 (0.72) Serbian: 4.66 (0.26)	-	Spanish: 6.84 (0.17) Swedish: 4.68 (0.41)	Spanish: 6.16 (1.24) Turkish: 4.38 (0.88)
	37; .76; 6.505e-08* 58%	10; NA	29; .72; 1.165e-05* 52%	45; .59; 2.26e-05* 35%	102; .57; 3.219e-10* 32%	27; .35; 0.07 12%	4; NA	4; NA	14; NA	13; NA	39; .59; 7.735e-05* 35%	7; NA		9; NA	20; .82; 6.018e-06* 67%
Swedish	Swedish: 6.57 (0.94) Basque: 6.48 (1.27)	Swedish: 2.23 (1.79) Catalan: 5.98 (1.38)	Swedish: 6.36 (0.95) Croatian: 4.63 (0.63)	Swedish: 5.65 (1.28) Eng_Bird: 5.02 (1.09)	Swedish: 6.09 (1.02) Eng_Cortese: 5.43 (1.36)	Swedish: 6.02 (1.1) English_MRC1: 5.18 (1.4)	Swedish 6.18 (1.23) Greek: 6.21 (1.24)	Swedish: 6.65 (0.38) Cypriot_Greek: 6.25 (0.93)	Swedish 4.93 (1.76) Hungarian: 5.45 (1.58)	Swedish: 5.32 (1.04) Italian: 5.81 (1.51)	Swedish: 5.97 (1.35) Norwegian: 5.79 (1.41)	Swedish: 6.6 (0.32) Serbian: 6.66 (0.64)	Swedish: 4.68 (0.41) Spanish: 6.84 (0.17)	-	Swedish: 6.04 (1.25) Turkish: 6.24 (1.23)
	31; .35; 0.05* 13%	18; NA	43; .69; 3.142e-07* 47%	27; .73; 1.224e-05* 53%	75; .74; 2.65e-14* 55%	28; .69; 4.549e-05* 48%	12; NA	12; NA	12; NA	19; NA	64; .78; 3.33e-14* 60%	12; NA	9; NA		24; .69; 6.218e-05* 48%
Turkish	Turkish;: 6.43 (0.86) Basque: 6.42 (1.32)	Turkish: 2.38 (1.67) Catalan: 5.64 (1.51)	Turkish: 4.42 (0.93) Croatian: 6.13 (1.18)	Turkish: 6.23 (0.86) Eng_Bird: 5.06 (1.12)	Turkish: 6.21 (1.09) Eng_Cortese: 5.85 (1.35)	Turkish: 6.25 (0.97) English_MRC1: 5.42 (1.23)	Turkish: 6.34 (1.01) Greek: 6.29 (0.87)	Turkish: 6.13 (1.16) Cypriot_Greek: 6.57 (0.61)	Turkish: 5.55 (1.43) Hungarian: 6.04 (1.4)	Turkish: 5.78 (0.45) Italian: 6.6 (0.6)	Turkish: 6.54 (0.63) Norwegian: 6.58 (0.69)	Turkish: 6.53 (0.44) Serbian: 6.64 (0.4)	Turkish: 4.38 (0.88) Spanish: 6.16 (1.24)	Turkish: 6.24 (1.23) Swedish: 6.04 (1.25)	-
	66; .53; 5.398e-06* 28%	28; -.92; 1.65e-11* 85%	83; .69; 2.171e-13* 47%	31; .31 0.08 9%	65; .62; 3.16e-08* 38%	45; .65; 1.59e-06* 43%	13; NA	15 NA	19; NA	30; .64; 9.799e-05* 41%	84; .30 0.005* 9%	19; NA	20; .82; 6.018e-06* 67%	24; .69; 6.218e-05* 48%	

Eng_Bird = data from Bird et al. (2001); Eng_Cortese = data from Cortese and Fugett (2004); Eng_MRCI1 = English data from Coltheart (1981); significant results at $p < .05$ are indicated with an asterisk. Non-significant correlations are highlighted in boldface.

Table S2. Number of equivalent words, p-value, rho, and R^2 for AoA scores

	Basque	Eng_ Bird	Eng_ Cortese	Norwegian	Italian	Spanish
Basque		83 0.002 -0.32 10%	192 1.306e-13 -0.5 25%	136 9.648e-10 -0.49 24%	43 0.012 -0.38 14%	209 4.599e-09 -0.39 15%
English_ Bird	83 0.002 -0.32 10%		95 2.2e-16 0.76 58%	294 2.2e-16 0.61 37%	124 3.453e-05 0.36 13%	929 2.2e-16 0.67 45%
English_ Cortese	192 1.306e-13 -0.5 25%	95 2.2e-16 0.76 58%		592 2.2e-16 0.69 48%	200 5.967e-09 0.39 15%	2263 2.2e-16 0.67 45%
Norwegian	136 9.648e-10 -0.49 24%	294 2.2e-16 0.61 37%	592 2.2e-16 0.69 48%		62 2.65-e-06 0.56 31%	549 2.2e-16 0.6 37%
Italian	43 0.012 -0.38 14%	124 3.453e-05 0.36 13%	200 5.967e-09 0.39 15%	62 2.65-e-06 0.56 31%		148 2.163e-15 0.59 35%
Spanish	209 4.599e-09 -0.39 15%	929 2.2e-16 0.67 45%	2263 2.2e-16 0.67 45%	549 2.2e-16 0.6 37%	148 2.163e-15 0.59 35%	

Basque = Duñabeitia, Casaponsa, Dimitropoulou, Martí, Larraza, & Carreiras (in prep.); Eng_Bird=Bird et al. (2001); Eng_Cortese=Cortese & Khanna (2008) for monosyllabic words + Schock, Cortese & Khanna (2012) for disyllabic words; semantically equivalent words. Norwegian ratings (Simonsen, Lind, Hansen, Holm, & Mevik, 2013; Lind, Simonsen, Hansen, Holm, & Mevik, 2015), Spanish ratings (Alonso, Fernandez, & Díez, 2015), Italian ratings (Rofes, de Aguiar, & Miceli, 2015).