Spanish affective normative data for 1,406 words rated by children and adolescents (SANDchild)

Lucía Sabater[1] · M. Guasch[2] · P. Ferré[2] · I. Fraga[3] · José A. Hinojosa[1,4,5]

Abstract

Most research on the relationship between emotion and language in children relies on the use of words whose affective properties have been assessed by adults. To overcome this limitation, in the current study we introduce SANDchild, the Spanish affective database for children. This dataset reports ratings in the valence and the arousal dimensions for a large corpus of 1406 Spanish words rated by a large sample of 1276 children and adolescents from four different age groups (7, 9, 11 and 13 years old). We observed high inter-rater reliabilities for both valence and arousal in the four age groups. However, some age differences were found. In this sense, ratings for both valence and arousal decreased with age. Furthermore, the youngest children consider more words to be positive than adolescents. We also found sex differences in valence scores since boys gave higher valence ratings than girls, while girls considered more words to be negative than boys. The norms provided in this database will allow us to further extend our knowledge on the acquisition, development and processing of emotional language from childhood to adolescence. The complete database can be downloaded from https://psico.fcep.urv.cat/exp/files/SANDchild.xlsx.

Keywords Emotional words · Children · Arousal · Valence · Affective database

Introduction

Language and emotion are intrinsically related from early childhood. Interestingly, this relationship emerges even before infants begin to produce words to denote feelings in the third year of life (Bahn, Vesker, García Alanis, Schwarzer, & Kauschke, 2017; Izard & Harris, 1995; Kristen, Sodian, Licata, Thoermer, & Poulin-Dubois, 2012). Children between

9 and 30 months produce fewer expressions with an emotional tone when they interact with their mothers in a playroom as they acquire complex language skills; apparently, older children with greater language abilities are able to express more emotions. This finding suggests that expressing emotions and learning words compete for limited cognitive resources in infants (Bloom, 1998). Later on, during the toddler period, children from 18 to 36 months use emotion labels to communicate their own or someone's else affective states and to manipulate other children's behaviors (Bretherton, Fritz, Zahn-Waxler, & Ridgeway, 1986). Three-year-old children raised in families in which conversations about feelings are relatively frequent became better by age 6 at judging emotions of unfamiliar adults (Dunn, Brown, & Beardsall, 1991).

Despite the existence of these strong connections, only a few studies have been concerned with the development of emotional language throughout childhood. These investigations can be grouped around two main issues. A first question relates to the acquisition, understanding, and use of emotional vocabulary based on either the reports provided by parents (Baron-Cohen, Golan, Wheelwright, Granader, & Hill, 2010; Bretherton & Beeghly, 1982; Li & Yu, 2015; Ridgeway, Waters, & Kuczaj II, 1985) or on a direct assessment of children's competence (Baron-Cohen et al., 2010). Overall, the results of these studies indicate that most children between 28 and 36 months know words describing basic

✉ Lucía Sabater
sabater93@gmail.com

✉ José A. Hinojosa
hinojosa@ucm.es

[1] Instituto Pluridisciplinar, Universidad Complutense de Madrid, Paseo Juan XXIII, 1, 28040 Madrid, Spain

[2] Department of Psychology and CRAMC, Universitat Rovira i Virgili, Tarragona, Spain

[3] Cognitive Processes & Behavior Research Group, Department of Social Psychology, Basic Psychology, and Methodology, Universidade de Santiago de Compostela, Santiago de Compostela, Spain

[4] Departamento Psicología Experimental, Procesos Cognitivos y Logopedia, Universidad Complutense de Madrid, Madrid, Spain

[5] Facultad de Lenguas y Educación, Universidad de Nebrija, Madrid, Spain

✎ Springer

feelings such as *scared*, *sad* or *happy* (Bretherton & Beeghly, 1982; Nook, Sasse, Lambert, McLaughlin, & Somerville, 2017; Ridgeway et al., 1985) and that the size of the emotional lexicon dramatically increases between 4 and 11 years of age (Baron-Cohen et al., 2010; Li & Yu, 2015). Also, it seems that children acquire words denoting positive concepts earlier in life than neutral or negative words (Baron-Cohen et al., 2010; Li & Yu, 2015).

A second line of research has focused on how children process emotional words at different ages. These studies are ultimately interested in establishing at what age children show sensitivity to emotional content in lexico-semantic processing. Using auditory lexical decision tasks (i.e., deciding whether a string of letters is a word; Bahn et al., 2017; Lund, Sidhu, & Pexman, 2019; Ponari, Norbury, & Vigliocco, 2018) and emotional categorization tasks (i.e., deciding whether a word denotes a positive, a negative or a neutral concept, both auditorily, e.g., Bahn et al., 2017, and visually, e.g., Sylvester, Braun, Schmidtke, & Jacobs, 2016), it was observed that 5–6-year-old children are able to derive significant benefit from emotional content in lexical processing (Bahn et al. 2017; Lund et al., 2019). Additionally, a processing advantage for positive words has been generally observed (Bahn et al., 2017; Lund et al., 2019; Ponari et al., 2018; Sylvester et al., 2016), which possibly arises from the earlier acquisition of positive compared to neutral or negative words during child development (Baron-Cohen et al., 2010; Li & Yu, 2015).

The small number of studies on the development of affective language may reflect, in part, difficulties in the selection of adequate stimuli. One possible reason is the absence of normative studies that provide norms for a word's emotional variables assessed by children of different ages. In fact, research on the interaction between language and emotion in adults has expanded in recent years as the availability of datasets with such types of ratings has increased (for reviews, see Citron, 2012; Fraga, Guasch, Haro, Padrón, & Ferré, 2018; Hinojosa, Moreno, & Ferré, 2019; Kissler, Assadollahi, & Herbert, 2006). Most of these normative studies are based on a dimensional theoretical approach of emotion (Russell & Ridgeway, 1983; Russell & Bullock, 1986; Scherer, 2005). According to this view, emotions can be defined in terms of at least two orthogonal dimensions, valence (ranging from pleasant to unpleasant) and arousal (ranging from activating to calming). Following the publication of the seminal ANEW database (Bradley & Lang, 1999), several normative studies have been conducted in different languages including English (Citron, Weekes, & Ferstl, 2014; Warriner, Kuperman, & Brysbaert, 2013), German (Schmidtke, Schröeder, Jacobs, & Conrad, 2014; Võ et al., 2009; Võ, Jacobs, & Conrad, 2006), French (Monnier & Syssau, 2014), Italian (Montefinese, Ambrosini, Fairfield, & Mammarella, 2014), Portuguese (Soares, Comesaña, Pihneiro, Simoes, & Frade, 2012), Dutch (Moors et al., 2013), Spanish (e.g., Guasch, Ferré, & Fraga, 2016; Hinojosa, Martínez-García, Villalba-García, Fernández-Folgueiras, Sánchez-Carmona, Pozo, & Montoro, 2016; Redondo, Fraga, Padrón, & Comesaña, 2007; Stadthagen-González, Imbault, Pérez-Sánchez, & Brysbaert, 2017), Chinese (Yao, Wu, Zhand, & Wang, 2017), Finish (e.g., Eilola & Havelka, 2010), Indonesian (Sianipar, van Groenestijn, & Dijkstra, 2016), Polish (Imbir, 2015) and Croatian (Ćoso, Guasch, Ferré, & Hinojosa, 2019). The existence of these databases has allowed a rapid increase in the knowledge about how adults process emotional language (see Citron, 2012; Hinojosa, Moreno, & Ferré, 2019). In contrast, normative studies reporting affective ratings for words by children are very scarce. To our knowledge, only five studies have collected scores for emotional properties of words from children or adolescents of different ages (Chinese: Ho, Mak, Yeung, Duan, Tang, Yeung & Ching, 2015; English: Vasa, Carlino, London, & Min, 2006; French: Monnier & Syssau, 2017; Syssau & Monnier, 2009; German: Sylvester et al., 2016). Notably, three of these studies included a rather small number of words (160 in Ho et al., 2015; 90 in Sylvester et al., 2016; 81 in Vasa et al., 2006), which were rated by only one group of participants (from 7 to 12 years old in Sylvester et al., 2016; from 12 to 17 years old in Ho et al., 2015) or by more than one group (three groups: 9, 10, and 11 year old raters in Vasa et al., 2006) and which were rated only in valence (Vasa et al., 2006) or both in valence and arousal (Ho et al., 2015; Sylvester et al., 2016). A larger corpus of words was assessed in two studies that collected valence scores for 600 words from 5, 7, and 9 year old French children (Syssau & Monnier, 2009), and valence and arousal ratings for 720 words from French children and adolescents of ages 7, 9, 11, and 13 (Monnier & Syssau, 2017).

The lack of normative studies reporting children's ratings for the emotional features of words is even more dramatic in the light of recent findings showing that there are important differences between adults and children in the way they perceive and process affective language. In particular, the development of verbal knowledge seems to mediate the expansion of the representation of emotional concepts and experiences from a positive-negative dichotomy at age 6 to a multidimensional organization in adulthood (Nook et al., 2017). Also, there is evidence indicating the existence of a positive bias in young children in comparison to adults. Indeed, 5-6 year-old children have shown better performance than adults with positive words in both a lexical decision task and an emotion categorization task (Bahn et al., 2017). These findings highlight the need to conduct normative studies including a high number of words rated by children and adolescents of different ages. This would allow researchers to overcome some limitations of prior studies on lexico-semantic processing in children which selected emotional words based on the ratings of adult participants (e.g., Lund et al., 2019; Ponari et al., 2018). Normative studies would provide researchers with stimuli

that are suited to the participant's age to conduct research on the acquisition and processing of emotional language in children. Therefore, in the current normative study we aim to collect ratings of valence and arousal for a large sample of 1406 Spanish words from four different age groups: 7-, 9-, 11-, and 13-year-old children. To our knowledge, this would be the largest published database reporting children's assessments of the emotional properties of words.

An additional goal of our study is to examine age-related effects and sex-related effects on the assessment of words' valence and arousal values. Prior findings indicate that younger children (7 and 9 years old) produce higher arousal and more positive scores than older children or adolescents (11 and 13 years old), and that boys give higher arousal ratings than girls (Monnier & Syssau, 2017).

Methods

Participants

One thousand two-hundred and seventy-six children and adolescents were recruited for the present study: 350 seven year-old children (173 girls, 177 boys; mean age = 7 years and 7 months, $SD = 4$ months), 318 9-year-old children (161 girls, 157 boys; mean age = 9 years and 6 months, $SD = 4$ months), 297 11-year-old children (157 girls, 140 boys; mean age = 11 years and 8 months, $SD = 4$ months), and 311 13-year-old children (155 girls, 156 boys; mean age = 13 years and 8 months, $SD = 4$ months). All were native Spanish speakers. The participants were recruited from several educational centers in the region of Madrid, including areas with different socioeconomic status. The study was approved by the local ethics committee. We obtained the consent from the management team of each educational center, as well as from the parents of each participant.

Materials and procedure

The Spanish affective normative data for children (SANDchild) consists of 1406 words. These words were selected from several adult databases which included affective ratings for Spanish words (Hinojosa et al., 2016a; Guasch et al. 2016; Redondo et al., 2007; Stadthagen-González et al., 2017). All the words had an age of acquisition under 7 years according to the scores of the normative studies of Alonso, Fernández, and Díez (2015) and Hinojosa et al. (2016b). The words were randomly distributed in 14 paper-and-pencil questionnaires of 70 words and 6 questionnaires of 71 words (one random order per questionnaire). Two versions of these 8-page questionnaires were made, one for valence and another for arousal. The first page included the personal information (first name and family name, sex, and date of birth), as well as one

example of a positive (*abuela, grandmother*), a negative (*coliflor, cauliflower*), and a neutral (*piedra, stone*) word. The remaining pages included 11 words printed in the center of the page with the exception of the last page, which contained 4 words (or 5 words in the questionnaires with 71 words). The corresponding valence (in valence questionnaires) or arousal (in arousal questionnaires) 9-point SAM scales (Self-Assessment Manikin; Lang, 1980) were printed under each word. The words were written in Times New Roman 22-point font in black and capital letters. Figure 1 shows word example from a valence and an arousal questionnaire.

Each child rated the valence and the arousal versions of the two questionnaires, with the exception of few 7-year-old participants that rated only one questionnaire. The 9-, 11-, and 13-year-old participants filled out the arousal questionnaires followed by the valence questionnaires at their own pace in groups, in a quiet room of their schools in a single session that lasted about one hour. Each participant fulfilled the valence and arousal versions of the same questionnaires. The 7-year-old group rated each version of the questionnaire in two separate sessions that lasted around 40 minutes each and which were separated between 1 to 10 days. Based on prior observations[1], in the first session we distributed the arousal questionnaires to each participant to avoid the children conflating arousal ratings with valence ratings. After participants filled in their personal information, the experimenter verbally described the arousal scale and explained to the children how they should use it to rate their feelings about the concepts denoted by the words. Once each child completed the arousal questionnaires, s/he waited until the whole class had finished. In the second session, the valence questionnaires were distributed and the experimenter verbally explained the valence scale and how to use it. For the 7-year-old children, the experimenter read aloud the examples of the negative, the positive and the neutral word in the first page. After asking children to verbally rate the examples, the experimenter gave feedback to children about their ratings. This procedure aimed to verify that children understood the instructions correctly and to show them that there were no correct and incorrect responses, so different answers were possible.

Results and discussion

Data trimming and description

Only fully completed questionnaires, in which the same children rated both valence and arousal for the same words, were

---

[1] We observed that when we distributed valence questionnaires in the first session, children gave low arousal scores to negative valence words in the second session, since these words had been previously scored low in the valence dimension. However, when we distributed the arousal questionnaires in the first session, children gave high arousal scores to negative words but they gave low valence ratings in the second session.
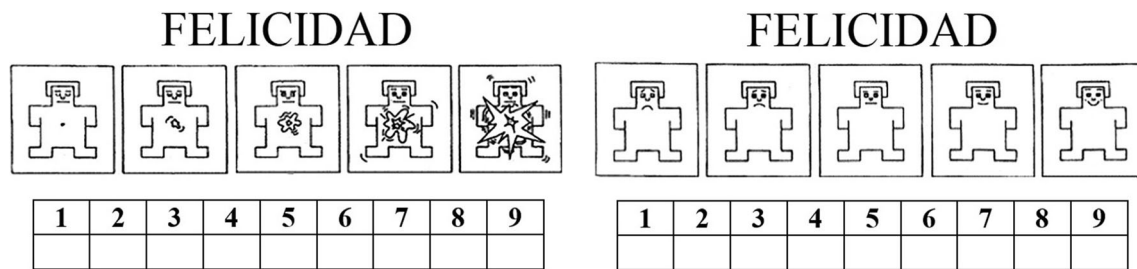
FELICIDAD

FELICIDAD

Fig. 1 Stimulus examples from a valence questionnaire and an arousal questionnaire showing a word (*felicidad*, happiness) with the SAM scales of arousal (left) and valence (right)

included in the study. Participants' responses were removed if they gave the same value to 90% or more of the words of a questionnaire. This led to the removal of thirteen 7-year-old participants, five 9-year-old participants and three 11-year-old participants (1.13% of the total). The total number of valid questionnaires (each questionnaire including ratings of valence and arousal for the same words) was 2280, evenly distributed by sex and age groups (see Table 1). Each word obtained a minimum of 25 ratings in each variable (M = 28.50, SD = 2.32, range [25-35]).

The complete database is available at https://psico.fcep.urv.cat/exp/files/SANDchild.xlsx. The file contains the 1406 Spanish words sorted in alphabetical order with their English translations, and 60 additional columns with all the data. The mean valence and arousal ratings for each word are included, as well as their standard deviations. This information is provided for the total sample. Additionally, separate data for girls and boys are provided. There is also a column indicating the total number of ratings for each word, and two more columns showing the number of boys and girls who rated that word. All this information is provided for each of the four age groups. Finally, we report data for a number of psycholinguistic variables from different databases when available. These variables include concreteness (taken from Duchon, Perea, Sebastián-Gallés, Martí, & Carreiras, 2013; Ferré, Guasch, Moldovan, & Sánchez-Casas, 2012; Guasch et al., 2016; Hinojosa et al., 2016a), familiarity (taken from Duchon et al., 2013; Ferré et al., 2012; Guasch et al., 2016), age of acquisition (taken from Alonso et al., 2015; Hinojosa et al., 2016b), word frequency at 7, 9 and 11 years (taken from Martínez & García, 2004), word frequency at adulthood (taken from Duchon et al., 2013) and number of orthographic neighbors (taken from Duchon et al., 2013).

Reliability and validity of the measures

To assess the inter-rater reliability of the valence and arousal ratings we computed the Intraclass Correlation Coefficients (ICCs) for each questionnaire. Then, the ICCs of the 20 versions were averaged for each variable and for each age group (see Table 2).

Both valence and arousal show high inter-rater reliabilities in the four age groups. Two findings are worth mentioning here. First, reliability tends to improve as age increases while variability (SD) decreases. Indeed, 7-year-old children show the lowest ICC values for both valence and arousal. Interestingly, looking at the mean valence (.97) and arousal (.88) ICCs for Spanish adult speakers reported in Guasch et al. (2016), it seems that reliability further increases until speakers reach adulthood. Second, ICCs are higher for valence than for arousal. This result suggests a higher consensus in valence than in arousal ratings, a finding that has been typically reported in normative studies in different languages and age groups (e.g., Guasch et al., 2016; Hinojosa et al., 2016a; Monnier & Syssau, 2017; Montefinese et al., 2014).

We also assessed the validity of our ratings. The most logical approach would have been to compare current scores to those reported by other normative studies with Spanish children. However, to the best of our knowledge, there are no such studies available. To overcome this limitation, we compared our data to those collected in other languages. In particular, we

Table 1 Distribution of raters across sex and age groups

N = 2280	7 years	9 years	11 years	13 years	Total
Boys	11.93%	12.68%	11.62%	13.55%	**49.78%**
Girls	11.58%	12.81%	12.37%	13.46%	**50.22%**
Total	**23.51%**	**25.48%**	**23.99%**	**27.02%**	**100.00%**

Table 2 Mean, standard deviation (SD), and range of the intraclass correlation coefficients of the questionnaires for each variable and age group

	Valence			
	7 years old	9 years old	11 years old	13 years old
Mean	.89	.94	.95	.94
SD	.04	.02	.01	.02
Range	[.76-.94]	[.90-.96]	[.92-.97]	[.91-.97]
	Arousal			
	7 years-old	9 years-old	11 years-old	13 years-old
Mean	.76	.80	.83	.85
SD	.08	.08	.06	.04
Range	[.65-.91]	[.62-.90]	[.69-.90]	[.79-.93]

decided to focus on the French study by Monnier and Syssau (2017). The reason was twofold. On the one hand, this study included the same age groups as ours. On the other, it is the normative study with the largest number of overlapping words with our database. We performed Pearson correlations with the 474 overlapping words. Correlation values for valence were $r(472) = .74$ for 7-year-old children, $r(472) = .76$ for 9-year-old children, $r(472) = .82$ for 11-year-old children, and $r(472) = .81$ for 13-year-old children (all $ps < .001$). Concerning arousal, correlations were $r(472) = .57$, $r(472) = .62$, $r(472) = .63$, and $r(472) = .56$ for 7-, 9-, 11- and 13-year-old children, respectively (all $ps < .001$). Thus, correlation values are rather high for valence and moderate for arousal. The higher validity of valence with respect to arousal is a common finding in adult affective ratings within and across languages (e.g., Eilola & Havelka, 2010; Guasch et al., 2016; Redondo et al., 2007; Soares, Comesaña, Pinheiro, Simoes, & Frade, 2012). However, our values are somewhat lower than those reported in such studies overall. For instance, the correlation coefficients for the comparison between the ANEW (Affective Norms for English Words; Bradley & Lang, 1999), and its Spanish adaptation (1034 overlapping words, Redondo et al., 2007) were .92 for valence and .75 for arousal. A possible explanation for these divergent findings is that child ratings might show higher variability than adult ratings. The results of the studies that compared ratings from different populations of children in the same language are in agreement with this idea. In this sense, Monnier and Syssau (2017) compared their valence ratings with those collected in a previous study, also with French children (Syssau & Monnier, 2009). The correlations were r = .79 and r = .83 for 7- and 9-year-old children, respectively. These values are very similar to those observed in the cross-language comparison between our ratings and those by Monnier and Syssau (2017). However, they are again lower than the correlations found for within-language comparisons in adults. For example, the comparison between the Guasch et al. (2016) and the Redondo et al. (2007) Spanish databases was .97 for valence and .84 for arousal. Hence, both within-language and cross-language comparisons reveal that child data are less consistent than adult data.

Age-related effects

First, we computed the Pearson correlations across age groups for valence and arousal. The correlation coefficients for valence showed similar positive and high values for all comparisons (see Table 3), which resemble previous findings from French children (Monnier & Syssau, 2017).

Concerning arousal (see Table 4), correlations were also high and positive, although we observed differences across ages. In particular, the correlations between the ratings from the 7-year-old group and the other age groups were lower than

Table 3 Correlations between valence ratings across age groups ($N = 1406$)

Age groups	7 years old	9 years old	11 years old	13 years old
7 years old	**.89**	.88	.87	.82
9 years old	.96	**.94**	.91	.88
11 years old	.95	.96	**.95**	.91
13 years old	.90	.94	.95	**.94**

All $ps < .001$. Values above the diagonal show the raw correlations. Following Brysbaert, Lagrou, and Stevens (2017), values below the diagonal show the disattenuated correlations obtained by using the reliability estimates in the main diagonal of the table

the correlations between the other three age groups, suggesting that the ratings by younger kids are the most divergent ones. Apart from that, correlations for arousal were lower than for valence. These findings are in line with the above mentioned reliability and validity measures, reinforcing the idea that arousal ratings are less consistent than valence ratings. They are also in line with the results of the few studies which have compared children and adult affective ratings (Bahn, Kauschke, Vesker, & Schwarzer, 2018; Russell & Paris, 1994; Sylvester, et al., 2016). In all these studies, arousal ratings are more weakly correlated between age groups and display higher variance than valence ratings.

Despite those high correlations, the visual inspection of the descriptive statistics suggested that there could be differences across ages in the absolute values for valence and arousal (see Table 5).

To examine age differences, we conducted two analyses of variance (ANOVA) with age group as a factor. The results for valence revealed a significant effect, $F(3, 4215) = 84.44$, $p < .001$, $\eta_p^2 = 0.06$, $MSE = 0.35$. Bonferroni-corrected comparisons showed that all the differences between age groups were significant (all $ps < .001$), except the comparison between the 7- ($M = 5.70$) and 11- ($M = 5.76$) year-old children ($p = .079$). Concerning arousal, the ANOVA again yielded a significant effect, $F(3, 4215) = 139.07$, $p < .001$, $\eta_p^2 = 0.09$, $MSE = 0.66$.

Table 4 Correlations between arousal ratings across age groups ($N = 1406$)

Age groups	7 years old	9 years old	11 years old	13 years old
7 years old	**.76**	.64	.56	.54
9 years old	.82	**.80**	.70	.65
11 years old	.71	.86	**.83**	.76
13 years old	.67	.79	.90	**.85**

All $ps < .001$. Values above the diagonal show the raw correlations. Following Brysbaert et al. (2017), values below the diagonal show the disattenuated correlations obtained by using the reliability estimates in the main diagonal of the table

Table 5 Valence and arousal descriptive statistics for each age and sex groups

	Valence											
	7 years old			9 years old			11 years old			13 years old		
	Boys	Girls	All	Boys	Girls	All	Boys	Girls	All	Boys	Girls	All
Mean	5.79	5.63	5.70	5.97	5.73	5.85	5.83	5.68	5.76	5.58	5.46	5.52
SD	1.56	1.89	1.60	1.71	1.88	1.72	1.64	1.81	1.67	1.38	1.56	1.42
Minimum	1.31	1.00	1.44	1.00	1.00	1.22	1.21	1.00	1.19	1.54	1.38	1.56
Maximum	9.00	9.00	9.00	9.00	9.00	8.97	9.00	9.00	8.96	8.71	8.88	8.72
	Arousal											
	7 years old			9 years old			11 years old			13 years old		
	Boys	Girls	All	Boys	Girls	All	Boys	Girls	All	Boys	Girls	All
Mean	5.48	5.27	5.38	5.46	5.30	5.40	5.18	5.38	5.27	4.80	4.95	4.88
SD	1.33	1.54	1.28	1.44	1.38	1.27	1.52	1.41	1.34	1.37	1.44	1.28
Minimum	1.75	1.57	2.33	1.33	1.55	1.83	1.46	1.46	1.74	1.07	1.47	1.88
Maximum	9.00	9.00	8.84	9.00	9.00	8.93	8.78	9.00	8.81	8.44	8.76	8.58

In this case, the only non-significant comparison was between the 7- (M = 5.38) and the 9- (M = 5.40) year-old children (p = 1). The remaining comparisons were significant (all ps < .006).

The above findings show a clear age effect for both valence and arousal: ratings for both variables tend to decrease as age increases. These results are in line with those reported by Monnier and Syssau (2017). These authors found higher ratings for valence and arousal in the younger children (they collapsed the 7- and 9-year-old groups into a single group) in comparison to the older children (the 11 and 13 year groups were also collapsed into a single group). Similarly, Bonivento, Tomasino, Garzitto, Piccin, Fabbro, & Brambilla (2017) reported higher valence (but not arousal) ratings in a group of 8-11-year-old participants, in comparison to a 12-15-year-old group. In contrast, such an age effect was not observed in the study conducted by Silvester and collaborators (2016) in German. Nonetheless, some caution is needed when interpreting the results of this last study since, as the authors themselves acknowledge, there was a small number of participants and words.

Our findings show differences in valence and arousal ratings across ages when all the words are taken together. However, we were also interested in examining whether such differences could be observed in all emotional categories or were they, rather, restricted to a particular type of word (i.e., positive, negative or neutral). To classify the words, we divided the 9-point Likert scale into three intervals of the same size. The intervals were as follows: Negative words were those located in the 1–3.66 valence range; neutral words in the 3.67–6.33 valence range and positive words in the 6.34–9 valence range (see Table 6).

Regardless of the age of the participants, the number of words that children considered as negative was lower compared to positive or neutral words (Table 6). The tendency of children to rate the words as more positive than adults seems

an unlikely explanation for these results. In this sense, the percentage of positive, negative and neutral words, with a lower proportion of negative words, is similar to those reported in normative studies in adults (e.g., Stadthagen-González et al., 2017). Additionally, data from Monnier and Syssau (2017), who also collected data from children, showed a similar distribution. Finally, there is evidence showing a negative correlation between age of acquisition and valence, so unpleasant words are learned later in life (e.g., Hinojosa et al., 2016b; Moors et al., 2013). Since all the words in our study have an age of acquisition under 7 years to assure that most participants knew their meaning (see also Monnier and Syssau, 2017, it is not striking that there are fewer negative words.

In order to examine age effects in valence and arousal ratings across emotional categories, we carried out a series of analyses. It should be noted that the particular set of words included in each range depended on participants' ratings, and for that reason they could not be exactly the same for the distinct age groups. Hence, it was not possible to perform an ANOVA here. Instead, we examined whether there were differences among age groups in the number of words considered as positive, negative and neutral. To that end, we carried out chi-squared tests for each type of word by taking the mean of the number of observations across the four age groups as the expected frequency (i.e., 529.25, 696.5, and 180.25 for

Table 6 Number of positive, neutral, and negative words in each age group

	7 years old	9 years old	11 years old	13 years old
Positive	522	627	564	404
Neutral	711	595	645	835
Negative	173	184	197	167

positive, neutral and negative words, respectively). Frequencies did not differ in the negative domain, $\chi^2(3) = 2.90$, $p = 0.407$. However, there were significant differences for positive words, $\chi^2(3) = 50.08$, $p < 0.001$, and for neutral words, $\chi^2(3) = 46.44$, $p < 0.001$. To identify the age groups in which those differences were significant, we carried out paired proportions tests. Concerning positive words, all the comparisons were significant (all ps < .015), with one exception, which was the comparison between 7- and 11-year-old children ($p = .104$). Regarding neutral words, all the comparisons across age groups were significant too (all ps < .013), except for the comparison between 9-year-old and 11-year-old children ($p = .058$). In sum, children of different ages do not greatly differ in the number of words considered as negative. However, if we focus on the two extreme groups (i.e., 7- and 13-year-old children), the youngest children consider more words as being positive and fewer words as being neutral compared to adolescents.

Sex-related effects

First, we computed the Pearson correlations between boys and girls across age groups (see Table 7). The data show that valence correlations are high, as in previous studies (Monnier & Syssau, 2017). Nonetheless, the correlations for arousal are again lower than those for valence. Also, these correlations increase with age, which suggests smaller sex differences as children grow up. In fact, 13-year-old boys and girls show correlations that are similar (e.g., Sianipar, Groenestijn, & Dijkstra, 2016) or even higher (e.g., Warriner, Kuperman, & Brysbaert, 2013) than those reported between sex groups in adult speakers.

In order to further examine sex differences in the absolute values for valence and arousal, we conducted two t-tests comparing ratings for boys and girls. The results suggested an effect for valence, $t = 8.67$, $p < .001$, indicating that boys gave higher valence ratings (M = 5.79) than girls (M = 5.62). No sex differences were observed for arousal (both M = 5.23, $t = 0.22$, $p = .826$). These results contrast with those of Monnier and Syssau (2017), who found a sex effect in arousal ratings (boys gave higher arousal values than girls), but not in valence ratings. They also differ from those of Sylvester et al. (2016), who reported higher valence values in girls in comparison to boys. In contrast, our results are in line with those reported in

some studies conducted with adult populations. Indeed, higher valence ratings have been reported for men than for women (e.g., Hinojosa et al., 2016; Montefinese et al., 2014), while no differences in arousal ratings between sexes have been found (Hinojosa et al., 2016; Redondo et al., 2007). It should be noted, however, that in other studies women gave higher arousal ratings than men (e.g., Söderholm, Häyry, Laine, & Karrasch, 2013; Soares et al., 2012). A possible reason for such discrepancies may be the different proportion of raters of each sex included in each study. The proportion of women included in adult studies is much larger than that of men. In contrast, studies conducted with children include a similar proportion of boys and girls. Therefore, sex comparisons should be more reliable in children than in adults. However, as stated above, differences in the sample size of participants and words across studies, as well as the lower consistency in children ratings (in comparison to adult ratings), may have contributed to the inconsistencies reported with children.

Finally, we were interested in knowing whether sex differences could be observed in all emotional categories or were, rather, restricted to a particular type of word (i.e., positive, negative or neutral). To that end, we computed the number of words considered as being positive, negative and neutral by boys and girls (see Table 8).

We carried out several proportions tests for independent measures. The results showed no significant sex differences in the number of positive words ($z = 0.27$, $p = 0.786$). However, there were differences for neutral words ($z = 2.26$, $p = 0.024$) and for negative words ($z = 3.75$, $p < .001$). Thus, it can be concluded that the differences in valence observed between boys and girls are mostly due to the fact that girls consider more words as being negative and fewer words as being neutral than boys. In order to address this issue in more detail, we identified the words that girls, but not boys, considered as negative, to explore whether these words could be related with a particular theme. To this end, we relied on the criterion used to classify the words as negative, neutral and positive explained above (i.e., negative words were those located in the 1–3.66 valence range). There were 74 words that the girls considered as negative and the boys considered as neutral or positive. We computed the difference in valence ratings between girls and boys for those words and identified those 25 words with the greatest score differences. A close inspection of these words suggest that they could be grouped

Table 7 Correlations for valence and arousal between boys and girls across age groups (N = 1406)

	7 years old	9 years old	11 years old	13 years old
Valence	.73	.81	.85	.87
Arousal	.59	.61	.68	.69

All ps < .001

Table 8 Number of positive, neutral, and negative words in each sex group

	Boys	Girls
Positive	538	531
Neutral	718	658
Negative	150	217

into two main different themes, one related to animals or insects (e.g., snake, spider or centipede), and another one related to weapons and violence (e.g., shot, gun or bullet).

The relationship between valence and arousal ratings

We examined the relationship between valence and arousal ratings across ages and sexes. To this end, we carried out a separate regression analysis for each age and sex group with valence as the independent measure and arousal as the dependent one.

The analyses focused on age revealed that, for 7-year-old children, there was a significant linear relation, $R = .79$, $F(1, 1404) = 2329.67$, $p < .001$, that accounted for 62.40% of the variance. A second-order polynomial fit was also significant, $R = .81$, $F(2, 1403) = 1381.13$, $p < .001$, accounting for 66.32% of the variance. Although the change in the variance explained by the second-order polynomial model in comparison to the linear model was modest (3.92%), it turned out to be significant ($p < .001$). Hence, the benefit of a second-order polynomial model for this age group was significant but small. Such benefit was larger for older children. Specifically, the relation between both variables was clearly nonlinear, $R = .74$, $F(2, 1403) = 845.67$, $p < .001$, in the 9-years-old group, where it explained 54.66% of the variance, in comparison to 35.01% of the variance explained by the linear relation. Similarly, the second-order relation, $R = .71$, $F(2, 1403) = 703.14$, $p < .001$, accounted for 50.06% of the variance (the linear relation accounted only for 16.34% of the variance) in the 11-year-old group, and for 48.78% of the variance, $R = .70$, $F(2, 1403) = 668.00$, $p < .001$ (in comparison to the 14.69% of the variance accounted for by the linear relation)

in the 13-year-old group. These results show that, as children get older, the pattern becomes more similar to that observed in adults, namely, the U-shaped relation between valence and arousal ratings in a two-dimensional affective space (Bradley & Lang, 1999; Eilola & Havelka, 2010; Ferré et al., 2012; Guasch et al., 2016; Hinojosa et al., 2016; Kanske & Kotz, 2010; Redondo et al., 2007; Soares et al., 2012; Võ et al., 2009).

A U-shaped relationship between valence and arousal means that the more affectively charged a word is (either in the positive or in the negative domain), the more arousing it tends to be. However, there seems to be an asymmetry between the negative and the positive poles in the ontogenetic development of the relation between valence and arousal. A visual inspection of Fig. 2 reveals that differences among the distinct age groups in arousal for the more positive words are very small. In contrast, the more negative words show substantial differences across ages. This pattern of findings suggests that negative words elicit a low degree of arousal in young children. As they grow up, these words would become more arousing.

In order to have a more complete picture of the effects of age on the relationship between valence and arousal, we identified the negative words which are not scored as arousing by 7-year-olds. To this end, we computed the average arousal value for negative words in the three older groups, and identified the words which are rated as arousing (in average) for the three older groups (arousal value higher than 5), but not for the 7-year-old children (arousal value below 5). There were 56 words that met these criteria. We computed the difference in arousal ratings between the younger group and the average of the three older groups and identified the 25 words showing the
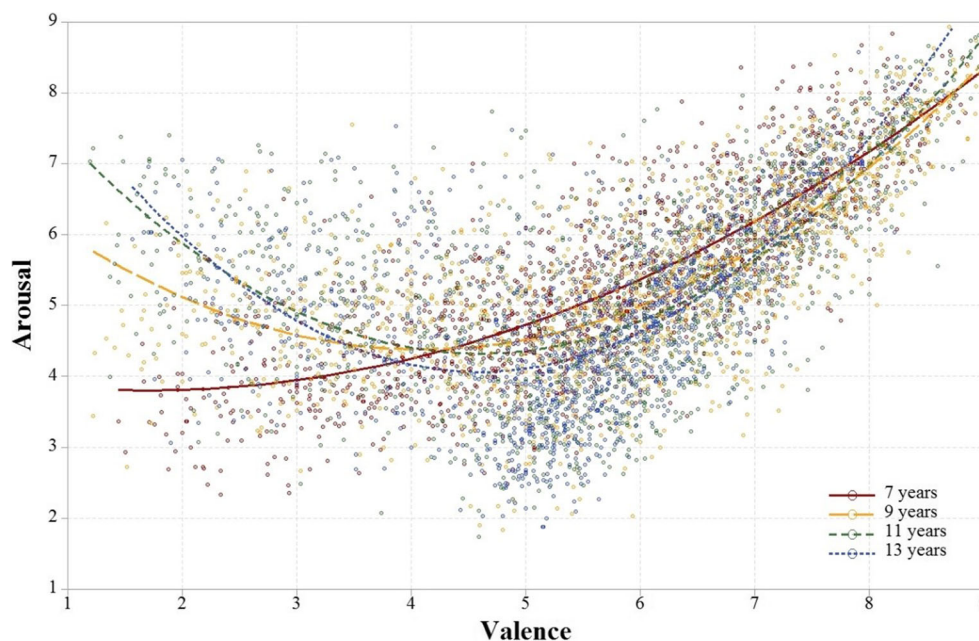

Fig. 2 Valence ratings plotted against arousal ratings for each age group (collapsing by sex)
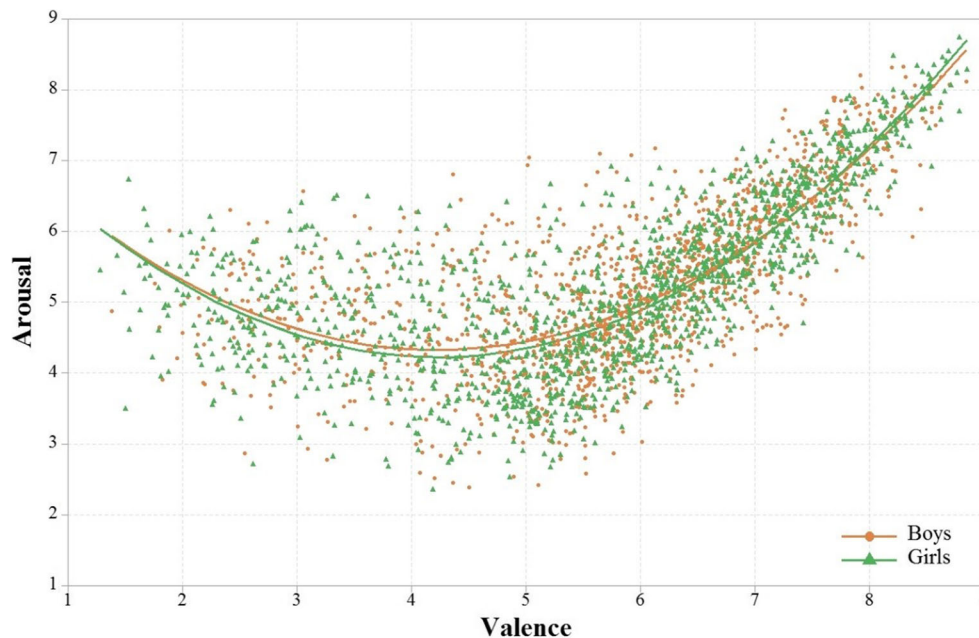
Fig. 3 Valence ratings plotted against arousal ratings for boys and girls (collapsing by age)

greatest score differences. Although a clear theme does not emerge, some of the words refer to bad objects, actions or situations (e.g., *kill, horror, suffocate, fire*). It may be that children, at these young ages, do not yet have strong experiences involving the reference of these words, so they assess these kinds of words as less activating. It is worth mentioning here the word showing the highest difference was *"fail"*, which is clearly arousing for the older groups, probably because they know the implications of failing an exam. In contrast, it was not arousing for 7-year-old children, who perhaps have not yet taken any exams.

Finally, the relationship between valence and arousal ratings across sexes was also examined. Here, the regression analysis for boys showed that the second-order relation was significant, $R = .75$, $F(2, 1403) = 908.51$, $p < .001$, accounting for 56.43% of the variance (the linear relation accounted for only 35.29% of the variance). A similar result was observed in girls, where the second-order relation was significant, $R = .82$, $F(2, 1403) = 1389.13$, $p < .001$, and explained 66.45% of the variance (the linear relation explained only 35.83% of the variance). Hence, as can be seen in Fig. 3, the pattern of relations between valence and arousal was very similar in boys and girls.

Conclusions

In this study a large group of children of different ages (7 years old, 9 years old, 11 years old and 13 years old) rated a large set of Spanish words in both the valence and arousal dimensions. Children's ratings (both in boys and girls) show the U-shaped relationship between valence and arousal typically observed in adults. The only exception is the group of 7-year-old children, who did not show high arousal ratings for the more negative words. The analyses of age-related effects reveal that both valence and arousal ratings tend to decrease as age increases. Regarding valence, it seems that changes in positive and neutral words account for this result, rather than changes in negative words. Indeed, the youngest children consider more words as being positive and fewer words as being neutral than the older children. With respect to sex-related effects, they are restricted to valence, where boys give higher valence ratings than girls. In this case, differences are found for negative and neutral words, because girls consider more words as being negative and fewer words as being neutral than boys. These results suggest that both age and sex differences in ratings should be taken into consideration when designing experiments aimed to study emotional word processing in children. This dataset will be very useful for researchers interested in this field, as it provides them with normative values for a large number of words. Hopefully, it will contribute to increasing the number of studies on the relation between language and emotion with a focus on children.

Appendix

The following texts are the instructions given to the participants, with their corresponding English translations. The instructions were the same for all group ages.

Instructions for the arousal dimension

Algunas veces, cuando leemos palabras, nos sentimos activados (la palabra "deporte", por ejemplo) o relajados (la palabra "masaje", por ejemplo). En este cuadernillo, vuestra tarea consistirá en indicar cómo de activados o relajados os hace sentir cada palabra. Imaginad que estáis escuchando música y ponemos el volumen a 1. Probablemente no os sentiréis muy activados. La música se escuchará más alto según vayamos incrementando el volumen. Imaginaos ahora que ponemos el volumen a 9, que es la potencia máxima. Ahora os sentiréis probablemente muy activados. Imaginad que leéis la palabra abuela. Si tuvierais que elegir un número entre 1 (me siento muy relajado) y 9 (me siento muy activado) para describir cuánta activación os hace sentir esta palabra, ¿qué número elegiríais? Recordad que cuanto más incrementemos el volumen, más activación sentiréis. ¿Y si leéis la palabra "coliflor"? Por ejemplo, a mí no me gusta la coliflor así que me hace sentir mucha activación. El volumen de mi activación es muy alto si leo esta palabra. Por último, ¿cómo os sentís con la palabra "piedra"?

Sometimes, when we read words, we feel activated (the word 'sport', for instance) or relaxed (the word 'massage', for instance). In this booklet, your task will be to indicate how much activation or relaxation each word makes you feel. Imagine that you are listening to music and the audio power is set at 1. You will possibly feel not so much activation. The music will be heard louder as we increase the audio power. Image now that we set the audio power at 9, which is the loudest power. Now you will possibly feel much activation. Imagine that you read the word grandmother. If you had to choose a number between 1 (I feel very relaxed) to 9 (I feel much activation) to describe how much activation this word makes you feel, which number would you choose? Remember that the louder we increase the power of the volume, the more activation you will feel. And if you read the word 'cauliflower'?. For example, I do not like cauliflower so it makes me feel much activation. The audio power of my activation is very high if I read to this word. Finally, how do you feel about the word 'stone'?.

Instructions for the valence dimension

Este cuadernillo contiene las mismas palabras que el que rellenasteis el otro día. Sin embargo, no debéis rellenarlo de la misma manera. Ahora distinguiremos entre emociones desagradables y agradables. Debéis marcar el número 1 cuando la palabra que leáis os parezca muy desagradable y el número 9 cuando os parezca muy agradable. Si la palabra que leáis os parece positiva, deberéis darle una puntuación entre 6 y 9. Cuanto más positiva os parezca, mayor puntuación. Por el contrario, si la palabra que leáis os parece negativa, tendréis que darle una puntuación entre 1 y 4. Cuanto menos os guste la palabra, más baja puntuación tendréis que darle. Además, hay algunas palabras que no nos hacen sentir ninguna emoción. En estos casos, deberíamos marcar el número 5. Por ejemplo, la palabra "abuela" me hace sentir bien, así que marcaré algún número entre el 6 y el 9. Sin embargo, no me gusta la "coliflor", así que marcaré algún número entre el 1 y el 4. Por último, la palabra "piedra" no me hace sentir ninguna emoción, así que marcaré el número 5.

This booklet has the same words as the one you filled the other day. However, you should not fill in the same way. Now we will have to differentiate between unpleasant and pleasant emotions. You should tick number 1 when the word you read seems very unpleasant to you and number 9 when it seems very pleasant to you. If the word you read seems positive to you, you will have to score between values 6 and 9. The more positive it seems to you, the higher the score. On the other hand, if the word you read seems negative to you, you will have to score between 1 and 4. The less you like the word, the lower score you will give to this word. Also, there are some words which make us feel no emotion at all. If this is the case, you should tick number 5. For example, the word 'grandmother' makes me feel good, so I will tick some number between 6 and 9. However, I do not like the word 'cauliflower', so I will tick a number between 1 and 4. Finally, the word 'stone' does not make me feel any emotion, so I will tick number 5.

References

Alonso, M. A., Fernandez, A., & Díez, E. (2015). Subjective age-of-acquisition norms for 7,039 Spanish words. *Behavior Research Methods, 47*, 268-274. https://doi.org/10.3758/s13428-014-0454-2

Bahn, D., Kauschke, C., Vesker, M., & Schwarzer, G. (2018). Perception of valence and arousal in German emotion terms: a comparison between 9-year-old children and adults. *Applied Psycholinguistics, 39*(3), 463-481. https://doi.org/10.1017/S0142716417000443

Bahn, D., Vesker, M., García Alanis, J. C., Schwarzer, G., & Kauschke, C. (2017). Age-dependent positivity-bias in children's processing of

emotion terms. *Frontiers in Psychology*, *8*, 1268. https://doi.org/10.3389/fpsyg.2017.01268

Baron-Cohen, S., Golan, O., Wheelwright, S., Granader, Y., & Hill, J. (2010). Emotion word comprehension from 4 to 16 years old: a developmental survey. *Frontiers in Evolutionary Neuroscience*, *2*, 109. https://doi.org/10.3389/fnevo.2010.00109

Bloom, L. (1998). Language development and emotional expression. New Perspectives in Early Emotional Development, 119.

Bonivento, C., Tomasino, B., Garzitto, M., Piccin, S., Fabbro, F., & Brambilla, P. (2017). Age-dependent changes of thinking about verbs. *Frontiers in Behavioral Neuroscience*, 11, 40. https://doi.org/10.3389/fnbeh.2017.00040

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

Bretherton, I., & Beeghly, M. (1982). Talking about internal states: The acquisition of an explicit theory of mind. *Developmental psychology*, *18(6):906-921*. https://doi.org/10.1037/0012-1649.18.6.906

Bretherton, I., Fritz, J., Zahn-Waxler, C., & Ridgeway, D. (1986). Learning to talk about emotions: A functionalist perspective. *Child development*, 57(3):529-548. https://doi.org/10.1111/j.1467-8624.1986.tb00225.x

Brysbaert, M., Lagrou, E., & Stevens, M. (2017). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism: Language and Cognition, 20*, 530–548. https://doi.org/10.1017/S1366728916000353

Citron, F. M. (2012). Neural correlates of written emotion word processing: A review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and Language, 122,* 211–226. https://doi.org/10.1016/j.bandl.2011.12.007

Citron, F. M., Weekes, B. S., & Ferstl, E. C. (2014). Arousal and emotional valence interact in written word recognition. *Language, Cognition and Neuroscience*, 29, 1257–1267. https://doi.org/10.1080/23273798.2014.897734

Ćoso, B., Guasch, M., Ferré, P., & Hinojosa, J. A. (2019). Affective and concreteness norms for 3,022 Croatian words. *Quarterly Journal of Experimental Psychology,* 72, 2302–2312. https://doi.org/10.1177/1747021819834226

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: one-stop shopping for Spanish word properties. *Behavior Research Methods, 45*, 1246-1258. https://doi.org/10.3758/s13428-013-0326-1

Dunn, J., Brown, J., & Beardsall, L. (1991). Family talk about feeling states and children's later understanding of others' emotions. *Developmental Psychology*, *27*(3), 448-455. https://doi.org/10.1037/0012-1649.27.3.448

Eilola, T. M., & Havelka, J. (2010). Affective norms for 210 British English and Finnish nouns. *Behavior Research Methods*, *42*(1), 134-140. https://doi.org/10.3758/BRM.42.1.134

Ferré, P., Guasch, M., Moldovan, C., & Sánchez-Casas, R. (2012). Affective norms for 380 Spanish words belonging to three different semantic categories. *Behavior Research Methods, 44*, 395-403. https://doi.org/10.3758/s13428-011-0165-x

Fraga, I., Guasch, M., Haro, J., Padrón, I., & Ferré, P. (2018). EmoFinder: The meeting point for Spanish emotional words. *Behavior Research Methods, 50*, 84-93. https://doi.org/10.3758/s13428-017-1006-3

Guasch, M., Ferré, P., & Fraga, I. (2016). Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behavior Research Methods, 48*, 1358-1369. https://doi.org/10.3758/s13428-015-0684-y

Hinojosa, J. A., Martínez-García, N., Villalba-García, C., Fernández-Folgueiras, U., Sánchez-Carmona, A., Pozo, M. A., & Montoro, P. R. (2016). Affective norms of 875 Spanish words for five discrete emotional categories and two emotional dimensions. *Behavior Research Methods, 48*, 272-284. https://doi.org/10.3758/s13428-015-0572-5

Hinojosa, J. A., Moreno, E. M., & Ferré, P. (2019). Affective neurolinguistics: towards a framework for reconciling language and emotion. *Language, Cognition and Neuroscience*, 1-27. https://doi.org/10.1080/23273798.2019.1620957

Ho, S. M., Mak, C. W., Yeung, D., Duan, W., Tang, S., Yeung, J. C., & Ching, R. (2015). Emotional valence, arousal, and threat ratings of 160 Chinese words among adolescents. *PloS One*, *10*(7), e0132294. https://doi.org/10.1371/journal.pone.0132294

Imbir, K. K. (2015). Affective norms for 1,586 Polish words (ANPW): Duality-of-mind approach. *Behavior Research Methods, 47*, 860–870. https://doi.org/10.3758/s13428-014-0509-4

Izard, C. E., & Harris, P. (1995). Emotional development and developmental psychopathology. In D. Cicchetti & D. J. Cohen (Eds.), *Wiley series on personality processes. Developmental psychopathology, Vol. 1. Theory and methods* (pp. 467-503). Oxford, England: John Wiley & Sons.

Kanske, P., & Kotz, S. A. (2010). Leipzig affective norms for German: A reliability study. *Behavior Research Methods, 42*(4), 987-991. https://doi.org/10.3758/BRM.42.4.987

Kissler, J., Assadollahi, R., & Herbert, C. (2006). Emotional and semantic networks in visual word processing: insights from ERP studies. *Progress in Brain Research, 156,* 147–183. https://doi.org/10.1016/S0079-6123(06)56008-X

Kristen, S., Sodian, B., Licata, M., Thoermer, C., & Poulin-Dubois, D. (2012), The Development of Internal State Language during the Third Year of Life: A Longitudinal Parent Report Study. *Infant and Child Development*, *21*(6): 634-645. https://doi.org/10.1002/icd.1767

Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications. In J. B. Sidowski, J. H. Johnson, & T. A. William (Eds.), *Technology in mental health care delivery systems* (pp. 119–137). Norwood, NJ: Ablex.

Li, Y., & Yu, D. (2015). Development of emotion word comprehension in Chinese children from 2 to 13 years old: Relationships with valence and empathy. *PloS One*, *10*(12), e0143712. https://doi.org/10.1371/journal.pone.0143712

Lund, T. C., Sidhu, D. M., & Pexman, P. M. (2019). Sensitivity to emotion information in children's lexical processing. *Cognition*, *190*, 61-71. https://doi.org/10.1016/j.cognition.2019.04.017

Martínez, J. A., & García, E. M. (2004). Diccionario frecuencias del castellano escrito en niños de 6 a 12 años. Servicio de Publicaciones Universidad Pontificia de Salamanca.

Monnier, C., & Syssau, A. (2014). Affective norms for French words (FAN). *Behavior Research Methods*, *46*(4), 1128-1137. https://doi.org/10.3758/s13428-013-0431-1

Monnier, C., & Syssau, A. (2017). Affective norms for 720 French words rated by children and adolescents (FANchild). *Behavior Research Methods*, 49, 1882-1893. https://doi.org/10.3758/s13428-016-0831-0

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2014). The adaptation of the affective norms for English words (ANEW) for Italian. *Behavior Research Methods, 46*, 887-903. https://doi.org/10.3758/s13428-013-0405-3

Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., Van Schie, K., Van Harmelen, A. L., … Brysbaert, M. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior Research Methods*, *45*(1), 169-177. https://doi.org/10.3758/s13428-012-0243-8

Nook, E. C., Sasse, S. F., Lambert, H. K., McLaughlin, K. A., & Somerville, L. H. (2017). Increasing verbal knowledge mediates development of multidimensional emotion representations. *Nature Human Behaviour*, *1*(12), 881-889. https://doi.org/10.1038/s41562-017-0238-7

Ponari, M., Norbury, C. F., & Vigliocco, G. (2018). Acquisition of abstract concepts is influenced by emotional valence. *Developmental Science*, *21*(2), e12549. https://doi.org/10.1111/desc.12549

Redondo, J., Fraga, I., Padrón, I., & Comesaña, M. (2007). The Spanish adaptation of ANEW (Affective Norms for English Words). *Behavior Research Methods, 39*, 600-605. https://doi.org/10.3758/bf03193031

Ridgeway, D., Waters, E., & Kuczaj II, S. A. (1985). Acquisition of emotion-descriptive language: Receptive and productive vocabulary norms for ages 18 months to 6 years. *Developmental Psychology, 21*, 901–908. https://doi.org/10.1037/0012-1649.21.5.901

Russell, J. A., & Paris, F. A. (1994). Do children acquire concepts for complex emotions abruptly? *International Journal of Behavioral Development, 17*(2), 349-365. https://doi.org/10.1177/016502549401700207

Russell, J. A., & Ridgeway, D. (1983). Dimensions underlying children's emotion concepts. *Developmental Psychology, 19*(6), 795-804. https://doi.org/10.1037/0012-1649.19.6.795

Russell, J. A., & Bullock, M. (1986). On the dimensions preschoolers use to interpret facial expressions of emotion. *Developmental Psychology, 22*(1), 97. https://doi.org/10.1037/0012-1649.22.1.97

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social science information, 44*(4), 695-729. https://doi.org/10.1177/0539018405058216

Schmidtke, D. S., Schroder, 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. https://doi.org/10.3758/s13428-013-0426-y

Sianipar, A., van Groenestijn, P., & Dijkstra, T. (2016). Affective meaning, concreteness, and subjective frequency norms for Indonesian words. *Frontiers in Psychology, 7*:1907. https://doi.org/10.3389/fpsyg.2016.01907

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*(1), 256-269. https://doi.org/10.3758/s13428-011-0131-7

Sylvester, T., Braun, M., Schmidtke, D., & Jacobs, A. M. (2016). The Berlin Affective Word List for Children (kidBAWL): Exploring processing of affective lexical semantics in the visual and auditory modalities. *Frontiers in Psychology, 7*, 969. https://doi.org/10.3389/fpsyg.2016.00969

Syssau, A., & Monnier, C. (2009). Children's emotional norms for 600 French words. *Behavior Research Methods, 41*(1), 213-219. https://doi.org/10.3758/BRM.41.1.213

Stadthagen-González, H., Imbault, C., Pérez Sánchez, M. A., & Brysbaert, M. (2017). Norms of valence and arousal for 14,031 Spanish words. *Behavior Research Methods, 49*, 111-123. https://doi.org/10.3758/s13428-015-0700-2

Söderholm, C., Häyry, E., Laine, M., & Karrasch, M. (2013). Valence and arousal ratings for 420 Finnish nouns by age and gender. *PloS ONE, 8*(8). https://doi.org/10.1371/journal.pone.0072859

Vasa, R. A., Carlino, A. R., London, K., & Min, C. (2006). Valence ratings of emotional and non-emotional words in children. *Personality and Individual Differences, 41*(6), 1169-1180. https://doi.org/10.1016/j.paid.2006.03.025

Võ, M. L., Jacobs, A. M., & Conrad, M. (2006). Cross-validating the Berlin affective word list. *Behavior Research Methods, 38*(4), 606-609. https://doi.org/10.3758/BF0319389

Võ, M. L., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin affective word list reloaded (BAWL-R). *Behavior Research Methods, 41*(2), 534-538. https://doi.org/10.3758/BRM.41.2.534

Warriner, A.B., Kuperman, V. & Brysbaert, M. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*, 1191–1207 (2013). https://doi.org/10.3758/s13428-012-0314-xhttps://doi.org/10.3758/s13428-012-0314-x

Yao, Z., Wu, J., Zhang, Y., & Wang, Z. (2017). Norms of valence, arousal, concreteness, familiarity, imageability, and context availability for 1,100 Chinese words. *Behavior Research Methods, 49*(4), 1374-1385. https://doi.org/10.3758/s13428-016-0793-2