

Perception of valence and arousal in German emotion terms: A comparison between 9-year-old children and adults

DANIELA BAHN and CHRISTINA KAUSCHKE
University of Marburg

MICHAEL VESKER and GUDRUN SCHWARZER
University of Gießen

Received: October 12, 2016 Accepted for publication: July 31, 2017

ADDRESS FOR CORRESPONDENCE

Daniela Bahn, Department of German Studies and Linguistics, University of Marburg, Marburg,
Pilgrimstein 16, 35037, Germany. E-mail: daniela.bahn@uni-marburg.de

ABSTRACT

Two major semantic features of emotion concepts have been shown to impact performance in emotion perception tasks: valence and arousal. To design psycholinguistic experiments with emotion terms as stimuli, norms are required that indicate valence and arousal values for individual words. Although such norms are usually obtained from ratings of adults, they are often also used in developmental studies. This procedure raises the question of whether children and adults perceive emotional valence and arousal of words in the same way, and consequently, whether adults' ratings are adequate when constructing stimulus sets for children. The present study obtained valence and arousal ratings for 48 German emotion terms from three different groups: 9-year-old children and adults tested in a controlled laboratory setting, and adults tested via online survey. Results demonstrate high correlations for valence and arousal across settings. The comparison between children and adults also revealed high correlations, suggesting that children at the age of 9 already display adultlike behavior in their evaluation of emotion terms. A small difference was found for absolute rating values of arousal, with children rating words less arousing than adults. Overall, 9-year-olds and adults are sufficiently similar in their perception of emotion to warrant the use of adult valence and arousal ratings in the analysis of children data.

Keywords: arousal; development; emotion terms; ratings; valence

The ability to share emotional information is highly relevant for social communication. While communicating, people have to express their own emotional states and infer those of others. Through language, emotions can be expressed by using different lexical means such as “affective words” and “emotion terms.” Emotion terms (e.g., “love”) refer to emotions as symbols (Schwarz-Friesel, 2013) and carry a specific value of emotionality (in comparison to neutral words).

© Cambridge University Press 2017 0142-7164/17

Furthermore, they are characterized by an individual degree of valence, which indicates the pleasantness of an emotional stimulus, that is, how positive or negative it is, and an individual degree of arousal, which refers to an emotion's intensity, that is, how strongly a person feels affected by it (reaching from high to low). Affective words, on the contrary, do not describe emotions but implicitly convey emotional information through connotations (e.g., "war"), which cause affective reactions. While affective words are either concrete (e.g., "cancer") or abstract (e.g., "burden"), emotion terms can contain both kind of information, since they are not as observable in the physical world as concrete words, but nevertheless physiologically perceivable. Thus, they can be seen as an independent category between abstract and concrete words (Vigliocco et al., 2013) whose acquisition builds a starting point into the acquisition of other abstract semantic representations.

Consequently, the acquisition of emotion terms constitutes an important part of lexical development. By participating in social and communicative situations, children gradually acquire emotion perception skills and establish conceptual knowledge about emotions. In their third year of life, children begin to verbally express emotional states through emotion terms (Bretherton & Beeghly, 1982). Studies in this field investigated at what age children have acquired concepts for certain emotions, and how they comprehend and produce emotion terms (e.g., Baron-Cohen, Golan, Wheelwright, Granader, & Hill, 2010; Kristen, Sodian, Licata, Thoermer, & Poulin-Dubois, 2012; Widen & Russell, 2010). The present study focuses on German children's categorization of emotion terms, in particular on their perception of valence and arousal as relevant semantic features of emotion terms.

VALENCE AND AROUSAL AS PARTS OF EMOTION CONCEPTS

Valence and arousal have been identified as two major dimensions along which different emotion concepts can be classified (Osgood, Suci, & Tannenbaum, 1957; Russell & Bullock, 1986; Russell & Ridgeway, 1983; Scherer, 2005).

Prior studies repeatedly found an impact of valence and arousal on adult word processing: emotionally toned stimuli clearly show a processing advantage over neutral stimuli (see Foolen, 2015, for a recent overview). However, it is unclear whether the polarity of emotional valence (i.e., whether the stimulus is positive or negative) also influences word processing, and if so, how.

As positive and negative objects of perception (including valenced words) have different meanings for the organism, it is natural that they are treated differently by the processing system in different tasks, and even in the language system itself. (Foolen, 2015, p. 247)

In addition, research is inconsistent with respect to the question of whether and how effects of valence and arousal might interact with each other (e.g., Hofmann, Kuchinke, Tamm, Võ, & Jacobs, 2009; Kuperman, Estes, Brysbaert, & Warriner, 2014). In short, valence and arousal crucially impact word processing. However, their influence has not been sufficiently explored yet, and may depend on factors such as social context, age, stimulus modality, or task.

Do the dimensions of valence and arousal also play a role in the development of emotion concepts? Russell and Ridgeway (1983) investigated children's unconscious knowledge of valence and arousal for English affective words using an emotional grouping task (sorting affective words by their similarities). It was found that children in the third grade already categorized the stimuli similarly to adults, but that they still became more mature (adultlike) in their categorizations as they grew older. Accordingly, valence and arousal are interpreted as salient components of emotional concepts that are unconsciously assessable by children, although they might not have fully acquired the emotional concept yet. According to Widen and Russell (2010), children first acquire two broad emotion categories based on the valence dimension (i.e., positive and negative) before these categories become more and more differentiated, and finally result in single emotion concepts with adultlike complexity. The reported results mentioned above suggest that the two dimensions of valence and arousal are already familiar to school-aged children, and that children intuitively use them to classify and categorize emotional stimuli.

VALENCE AND AROUSAL NORMS FOR AFFECTIVE AND EMOTIONAL WORDS

Given that valence and arousal influence the processing of affective words and emotion terms, stimuli used in experiments should be controlled for these two dimensions in order to get reliable and valid results (Citron, Weekes, & Ferstl, 2014; Schmidtke, Schröder, Jacobs, & Conrad, 2014).

Whereas many studies collected adults' norms of affective words that sometimes also include emotion terms (e.g., for German: Vö et al., 2009, for English: Bradley & Lang, 1999; Janschewitz, 2008, for Dutch: Moors et al., 2013, for Polish: Imbir, 2016), only a few studies so far focused on absolute values of valence and arousal estimates given by children. They are limited to a small number of words and only exist for a few languages. Therefore, adults' norms are often used for stimulus selection in developmental studies, because the assessment of children-based norms is costly and time consuming. The question arises whether this common approach generates item sets that are adequate for experiments with children. Do children differ from adults in the way they perceive and judge emotional information in words with regard to valence and arousal? If so, the use of adult norms for stimulus selection in developmental studies would be questionable. In the following, we will review existing evidence for age effects in studies assessing the judgments of valence and arousal for words.

Syssau and Monnier (2009) conducted a rating study with children aged 5, 7, and 9 years including 600 positive, negative, and neutral French words. The authors found that younger and older children differed in how they judge a word's valence. With increasing age, children rated words more often as neutral rather than positive, while the number of negatively rated words remained stable across the age groups. Vasa, Carlino, London, and Min (2006) collected valence norms of 81 English positive, neutral, and threat words, obtained from 9- to 11-year-old children. Results showed no age-related differences, indicating that children were able to differentiate the three categories with a strong internal consistency. Three

rating studies of valence and arousal with young, middle-aged, and older adults (French affective words: Gilet, Grünh, Studer, & Labouvie-Vief, 2012; English affective words: Grunwald et al., 1999; Finnish affective words: Söderholm, Häyry, Laine, & Karrasch, 2013) reported the existence of changes in adults' perception of valence and arousal over the entire life span. The three studies agree that older participants perceive affective words (regardless of the language) as more arousing and more positive than younger participants.

Only two studies so far compared children's and adults' norms of valence and arousal: Russell and Paris (1994) collected ratings for eight English emotion terms (two basic emotions and six complex emotions) from adults and 4- to 5-year-old children. The children were told that a stick figure feels a certain way (representing one of the eight emotions) and then asked to assign it to a 5-point Likert scale of valence and arousal. Adults conducted a simple rating task of both dimensions with the same scales. Sylvester, Braun, Schmidtke, and Jacobs (2016; pilot study reported in Jacobs et al., 2015) collected ratings from adults and 7- to 12-year-old children for 90 positive, negative, and neutral German words selected from the Berlin Affective Word List Reloaded (BAWL-R; Vö et al., 2009). Using a 5-point Likert scale for valence as well as for arousal, the children were asked to indicate how pleasant/unpleasant and calm/exciting the words feel to them. In both studies, high correlations were found suggesting that adults' norms serve as a significant predictor for children's ratings of both valence and arousal. Children and adults showed the same judgment behavior, indicating that at the age of 4 years, children have already acquired a basic understanding of emotion concepts. In addition, in both studies arousal ratings were more weakly correlated, and displayed a higher variance than the valence ratings.

To summarize, studies that systematically compared adults' and children's perception of valence and arousal in words are sparse so far. Independent from the applied methodology (grouping or rating task), the reported studies agree that children seem to perceive the dimension of emotional valence and, to a lesser extent, the dimension of arousal in a comparable way to adults. However, these studies mainly used a mixed set of neutral words, affective words, and emotion terms with a varying value of concreteness (Sylvester et al., 2016) or investigated a very limited number of emotion terms (Russell & Paris, 1994).

OBJECTIVE

The present study aimed at investigating children's and adults' evaluations of 48 German emotion terms along the dimensions of valence and arousal. The stimulus set was restricted to emotion words (excluding affective words) in order to avoid confounding effects of concreteness, which has been shown to impact word processing tasks (Kanske & Kotz, 2007). As already mentioned, affective words can be either concrete or abstract, whereas emotion terms contain both types of information. Therefore, affective words vary more with respect to their level of concreteness as opposed to emotion terms. The stimuli were controlled for a number of linguistic variables. Ratings of children and adults were compared using correlational analyses and group comparisons of absolute values.

In line with previous findings, we expected significant positive correlations for arousal and valence ratings obtained from 9-year-old children and adults. In addition, we expected that the correlation for arousal might be weaker than the correlation for valence (as already found by Russell & Paris, 1994; Schmidtke et al., 2014; Sylvester et al., 2016). Concerning the question of whether children differ from adults regarding the absolute values for valence and arousal, previous research is ambiguous and does not support a clear hypothesis. On the one hand, similarities were found in emotional grouping tasks (e.g., Russell & Ridgeway, 1983; Widen & Russell, 2010) as well as in the rating studies by Sylvester et al. (2016) and Russell and Paris (1994). On the other hand, some studies suggest developmental changes with respect to the perception of valence and arousal (Russell & Ridgeway, 1983; Widen & Russell, 2010) in children as well as age-related differences between young, middle-aged, and older adults (Gilet et al., 2012; Söderholm et al., 2013).

METHOD

Two computer-based rating tasks were conducted in order to collect judgments of valence and arousal for a set of 48 positive and negative German emotion terms. First, adult rating values collected via online survey will be compared to adult ratings obtained from the BAWL-R (Võ et al., 2009) for the same words. Second, in order to investigate age-related differences, the ratings of children and another adult group (collected in the laboratory under the same experimental settings) will be compared.

Participants

For the online rating task with adult participants, all students and employees of the Universities of Giessen and Marburg (Hesse, Germany) received an invitation by e-mail. In all, data were analyzed from 337 German native speakers (275 women, $M = 31$, years, 7 months, $SD = 12$ years, 6 months) for the arousal rating task and from 261 German native speakers (198 women, $M = 30$ years, 5 months, $SD = 11$ years, 4 months) for the valence rating task. The rating values were compared to the values reported in the BAWL-R studies, which were obtained from about 200 participants.

For the laboratory experiments, 30 typically developing 9-year-old children (15 girls, $M = 9$ years, 5 months, $SD = 3$ months) participated in the arousal rating and another 30 children in the valence rating (16 girls, $M = 9$ years, 9 months, $SD = 5$ months). All children were recruited in Giessen, Marburg, and surrounding, and grew up as monolingual native speakers of German. The existence of any developmental disorders had to be negated; corrected to normal vision was accepted. In order to confirm this, parents filled out a questionnaire on their children's development. Participants were randomly assigned to either the valence or the arousal group. In addition, 60 adults (employees and students from the University of Marburg and Giessen) participated in the laboratory experiment (30 in each rating) under the same conditions as the children (valence rating: 25 women, $M = 37$ years, 8 months, $SD = 42$ years; arousal rating: 27 women, $M = 27$ years, 8 months, $SD = 7$ years, 8 months).

Stimuli

The presented stimuli consisted of a set of 24 positive and 24 negative German emotion terms (see Table 1) such as “Glück” (Engl. *luck*), “leiden” (Engl. *to suffer*), or “tapfer” (Engl. *brave*) controlled for arousal and six linguistic variables. Stimulus construction comprised the following steps:

1. For a preselection of stimuli, 106 emotion terms (including negative and positive ones) were chosen from the BAWL-R. The BAWL-R database offers 3,000 German affective words (including emotion terms) with values for valence and arousal, rated by 200 adults (Võ et al., 2009). The BAWL-R also offers norms for several linguistic variables that impact on performance in word perception tasks: imageability, word class, number of letters, phonemes and syllables, word frequency, number and frequency of orthographic neighbors, and bigram frequencies.
2. In the second step, we defined seven linguistic variables, for which the item set should be controlled: number of phonemes and morphemes, neighborhood density, frequency, word class, age of acquisition (AoA), and concreteness. Norms that were not offered by the BAWL-R were gathered from other databases.
 - The number of phonemes is as given in the BAWL-R.
 - Morphological complexity was operationalized in the number of morphemes and determined for each of the 106 emotion terms.
 - Neighborhood density is as indicated in the BAWL-R, operationalized in neighbors at Edit Distance 1 (number of words that can be created from one particular word by changing one single letter).
 - *Frequency*: Because the selected words were to be rated by children, the frequency values offered by the BAWL-R, which are based on the CELEX database (Baayen, Gulikers, & Piepenbrock, 1995), were considered not to be appropriate for experiments with children. The CELEX frequencies are mainly based on written newspaper articles. Instead, the ChildLex corpus (German Children’s Book Corpus; Schröder, Würzner, Heister, Geyken, & Kliegl, 2015) was applied. This database consists of linguistic norms (word frequency, word length, and orthographic neighborhood size) for 10 million words, extracted from children’s books.
 - *AoA*: In order to gain AoA norms for the selected emotion words, a rating via online survey with employees and students of the Universities of Marburg and Giessen was conducted. Participants estimated on a 7-point Likert scale (from the age of 2 to 8 years and older) at what particular age a child most probably knows the meaning of each of the emotion terms. AoA values were derived from the means of all responses.
 - *Concreteness*: Norms of concreteness for the two sets of emotion terms were collected from 411 participants (employees and students) in another online rating study at the Universities of Marburg and Giessen. Using a 7-point Likert scale from 1 (*very abstract*) to 7 (*very concrete*), participants were asked to assign a specific value of concreteness to each of the emotion terms.
3. In the third step, two sets of emotion words (positive and negative ones according to the valence values from the BAWL-R) were created that clearly differ with

Table 1. Selected word sets of positive and negative emotion terms

Emotion Term	Word Class	English	Valence	Valence				Arousal			
				Children Lab	Adults Lab	Adults Online	Adults BAWL	Children Lab	Adults Lab	Adults Online	Adults BAWL
Trotz	Noun	Defiance	Negative	3.20	2.92	2.95	3.18	2.56	2.83	2.57	3
Sorgen	Noun	Worries	Negative	2.70	2.28	2.18	3.38	3.63	3.27	3.50	3.05
Zweifel	Noun	Doubt	Negative	2.87	2.56	2.71	2.50	2.83	3.30	3.02	3.10
Gejammer	Noun	Lamentation	Negative	2.17	2.32	2.00	2.30	2.83	2.93	2.76	3.22
Pech	Noun	Misfortune	Negative	1.97	2.32	2.06	2.00	2.47	2.73	2.49	3.33
Neid	Noun	Envy	Negative	3.11	2.16	1.92	2.10	3.08	3.43	2.71	3.56
Schuld	Noun	Fault/blame	Negative	2.53	2.28	1.67	2.30	3.20	3.50	3.43	3.57
Angst	Noun	Fear	Negative	2.00	2.00	1.79	1.40	3.10	4.10	3.73	4.38
Übermut	Noun	Wantonness	Negative	4.36	3.84	3.73	3.90	3.07	3.50	2.49	3.72
Aggression	Noun	Aggression	Negative	2.15	1.80	1.89	2.24	3.74	4.50	3.72	4.33
bereuen	Verb	To regret	Negative	2.53	2.76	2.76	3.40	3.00	3.17	2.98	2.79
trauern	Verb	To grieve	Negative	1.60	1.60	2.16	2.00	3.03	3.73	3.73	3.06
schreien	Verb	To shout	Negative	2.43	3.08	2.34	2.76	3.33	4.40	3.06	3.71
weinen	Verb	To cry	Negative	1.50	2.36	2.53	3.12	2.93	3.27	3.34	3.78
leiden	Verb	To suffer	Negative	1.83	1.48	1.44	2.06	3.03	3.70	3.57	3.83
heulen	Verb	To wail	Negative	1.20	2.00	1.95	2.30	3.00	3.83	3.01	3.85
fürchten	Verb	To be frightened	Negative	2.17	2.16	2.23	2.35	3.43	3.60	3.28	3.86
Schocken	Verb	To shock	Negative	2.43	2.56	2.51	2.60	3.29	4.13	3.31	3.96
Verletzen	Verb	To hurt sb.	Negative	2.07	2.08	1.77	2.06	2.93	4.03	3.54	4.03
Mutlos	Adjective	Discouraged	Negative	2.33	2.16	2.13	2.60	2.60	2.33	2.43	2.00
Lustlos	Adjective	Unenthusiastic	Negative	3.00	2.44	2.42	2.50	2.07	1.80	1.95	2.28
Ruhelos	Adjective	Restless	Negative	3.17	2.88	2.93	2.60	3.00	3.00	2.66	2.58
Schwach	Adjective	Weak	Negative	2.60	3.12	2.87	2.40	2.10	1.83	2.40	2.61
Sauer	Adjective	Sore/sour	Negative	2.33	2.52	2.65	3.10	3.40	3.73	3.09	3.53
Dank	Noun	Thanks/gratitude	Positive	5.80	5.60	5.99	6.53	2.53	2.47	2.80	2.28

Table 1 (cont.)

Emotion Term	Word Class	English	Valence	Valence				Arousal			
				Children Lab	Adults Lab	Adults Online	Adults BAWL	Children Lab	Adults Lab	Adults Online	Adults BAWL
Sicherheit	Noun	Safety	Positive	5.77	5.68	5.77	6.43	2.43	2.30	2.87	2.89
Glück	Noun	Luck	Positive	6.53	6.28	6.43	6.87	2.87	3.60	3.52	2.93
Interesse	Noun	Interest	Positive	5.37	5.48	5.52	6.80	2.80	2.80	2.42	3.19
Stolz	Noun	Pride	Positive	6.03	5.00	4.72	6.80	2.80	3.43	2.87	3.29
Neugier	Noun	Curiosity	Positive	5.10	5.56	5.30	7.70	3.70	3.30	2.91	3.56
Spaß	Noun	Fun	Positive	6.57	5.76	6.35	6.87	2.87	3.83	3.21	3.58
Lust	Noun	Pleasure	Positive	5.67	5.52	5.75	6.23	2.23	3.60	3.37	3.67
Mut	Noun	Courage	Positive	5.80	5.76	5.84	6.93	2.93	2.97	3.23	3.67
Freiheit	Noun	Freedom	Positive	6.33	6.00	6.47	7.43	3.43	3.53	3.63	4.00
Hoffen	Verb	To hope	Positive	4.60	5.44	5.24	6.90	2.90	2.90	3.00	2.39
Lachen	Verb	To laugh	Positive	6.57	6.24	6.44	6.93	2.93	3.53	3.43	2.83
Kichern	Verb	To giggle	Positive	5.70	5.28	5.09	6.50	2.50	2.70	2.25	2.96
Staunen	Verb	To gape	Positive	5.53	5.12	5.23	6.83	2.83	3.23	2.64	3.11
Mögen	Verb	To like	Positive	6.00	5.32	5.51	6.63	2.63	2.43	2.61	3.16
Zujubeln	Verb	To hail/cheer sb.	Positive	6.40	5.52	5.32	7.03	3.03	3.93	2.78	3.17
Wünschen	Verb	To wish	Positive	5.40	5.32	5.09	6.90	2.90	2.77	2.53	3.26
Begeistern	Verb	To inspire	Positive	6.03	6.00	5.99	7.13	3.13	3.97	3.36	3.50
Lieben	Verb	To love	Positive	6.40	6.40	6.60	7.10	3.10	4.20	3.94	3.67
Vertraut	Adjective	Familiar	Positive	5.53	6.00	5.89	6.23	2.23	2.33	2.91	1.61
Froh	Adjective	Glad/cheerful	Positive	6.00	5.84	5.85	6.80	2.80	2.97	2.75	2.28
Freudig	Adjective	Joyful	Positive	6.33	5.80	6.13	7.29	3.29	3.27	2.76	2.94
Tapfer	Adjective	Valiant/brave	Positive	5.69	5.36	5.36	7.07	3.07	2.53	2.68	3.11
Stark	Adjective	Strong	Positive	5.70	5.36	5.53	7.07	3.07	3.00	2.62	3.58

Note: BAWL-R, Affective Word List Reloaded.

Table 2. *Matching of positive and negative emotion terms*

	Average Values		<i>p</i>	Partial η^2
	Negative Words	Positive Words		
Valence	2.54	5.75	.000	0.850
Arousal	3.38	3.1	.116	0.053
Absolute frequency (1/mio)	40.74	62.49	.477	0.027
Number of morphemes	1.71	1.71	1.000	0.000
Number of phonemes	5.25	5.38	.943	0.003
Age of acquisition (years;month)	5;1	4;7	.483	0.029
Neighbors at edit distance 1	7.21	7.54	.986	0.056
Concreteness	3.77	3.72	.809	0.001

respect to valence, but are controlled for arousal, frequency, AoA, concreteness, word class, morphological complexity, word length, and neighborhood density. After the final matching procedure, each set comprised 24 items (10 nouns, 9 verbs, and 5 adjectives in each set). One-way analyses of variance confirmed that the two sets differed significantly with respect to valence, but not with respect to any other variable (see Table 2).

Finally, all 48 emotion terms were recorded in a soundproof booth, spoken by trained native speakers of standard German (one male and one female). Both speakers were instructed to use neutral prosody. In order to avoid gender biases, half of the participants in each rating group heard the words spoken by the male speaker, and the other half of them heard the words spoken by the female speaker.

Procedure

The rating experiment with children and adults took place at university laboratories in Marburg or Giessen. Prior to the experiment, parents were informed about the procedure and signed informed consent documents. Participants were comfortably seated in a quiet room, provided with headphones and with a 15.6-inch computer screen in front of them. A verbal instruction was given; either the Self-Assessment Manikin (SAM) scale by Bradley and Lang (1994) for valence (7-point Likert scale) or the 5-point SAM Likert scale for arousal was explained. On the valence scale, three manikins on the left side of the neutral middle (named 1, 2, and 3) indicated negative words and the three manikins on the right side (number 5, 6 and 7) indicated positive words. Children of the valence group were asked to estimate how positive (“good”/“pleasant”) or negative (“bad”/“unpleasant”) the meaning of each word is. The scale was explained with two example words with contrasting valence.

While “to laugh” was said to be positive, which is why a manikin from the right side needs to be chosen, for the negative word “angry” a manikin from the left side

would be appropriate. In addition, the explanation was given that some words might be more or less positive or negative than other words and that therefore the three manikins on each side of the scale show a different intensity of (un)pleasantness. The neutral manikin in the middle was told to be chosen whenever the participant thinks that a word is neither clearly positive nor negative. Children in the arousal group were supposed to judge how arousing they perceive the meaning of the words by choosing one of five different manikins. The first manikin on the left of the scale (number 1) represented very low arousing words (as an example the word “calm” was used), while manikin number 5 (on the right) represented words with a very high level of arousal (exemplified by the word “love”). Moreover, there were options to indicate “play the word again” and “I do not know the word.” Each experimental trial started with a black screen and the auditorily presented stimulus. After each word was heard, the rating scale and the two additional buttons appeared on the computer screen and stayed until the participant indicated his or her decision by a mouse click. Words were presented in randomized order. Both ratings were preceded by a training phase of 10 trials. The duration of the rating task was approximately 10 min per child. Each child was rewarded with a small gift at the completion of the experiment.

For adult participants, the experimental procedure took between 5 and 8 min and was the same as for the children, except of that no paraphrasing terms (such as “good” or “bad”) were used to further explain the concepts positive and negative. Prior to their participation, adults had to sign informed consent documents. They were either monetarily rewarded or received student credit for their effort.

For the online survey, the same stimuli, scales, and instructions were used with the only difference that words were not presented auditorily, but in written format. The experiments were programmed using OpenSesame (Mathôt, Schreij, & Theeuwes, 2012) as the controlling software.

Data analysis

The average rating values for valence and arousal for each word were calculated across participants (four different sample types: adult ratings according to BAWL-R, adult ratings via online study, adult ratings via laboratory setting, and child ratings via laboratory setting) and tested for Spearman correlations. Absolute values were then further analyzed using a separate linear mixed model (LMM) for valence and arousal, performed to investigate main effects of age and valence, as well as an interaction of these two factors. Words and subjects were treated as random factors, while valence (positive vs. negative) and age (children vs. adults) were fixed factors. Statistical significance was set at the 95% confidence level ($p < .05$). Commonly used statistical methods to compare averaged values of different samples, such as analyses of variances treat participants and items (e.g., words) as fixed factors and therefore disregard their variability (Baayen, Davidson, & Bates, 2008). The LMM avoids this so-called language as fixed effect fallacy by classifying participants and items as random factors. There are two more advantages of LMMs: no information gets lost by averaging across subjects or items, since raw data is used, and the LMM accounts for missing observations in the data set. The LMM analyses were performed with SPSS Version 22.

Table 3. *Correlations of valence ratings*

	Child Ratings			Laboratory			Adult Ratings Online Survey			BAWL		
	All Words	Positive Words	Negative Words	All Words	Positive Words	Negative Words	All Words	Positive Words	Negative Words	All Words	Positive Words	Negative Words
Child r_S Ratings				.905**	.546**	.689**	.905**	.651**	.586**	.897**	.631**	.560**
<i>p</i>				.000	.006	.000	.000	.001	.003	.000	.001	.004
<i>N</i>				48	24	24	48	24	24	48	24	24
Adult r_S Ratings lab	.905**	.546**	.689**				.971**	.914**	.854**	.941**	.801**	.729**
<i>p</i>	.000	.006	.000				.000	.000	.000	.000	.000	.000
<i>N</i>	48	24	24				48	24	24	48	24	24
Adult r_S Ratings online survey	.905**	.651**	.586**	.971**	.914**	.854**				.960**	.909**	.771**
<i>p</i>	.000	.001	.003	.000	.000	.000				.000	.000	.000
<i>N</i>	48	24	24	48	24	24				48	24	24
Adult r_S Ratings BAWL	.897**	.631**	.560**	.941**	.801**	.729**	.960**	.909**	.771**			
<i>p</i>	.000	.001	.004	.000	.000	.000	.000	.000	.000			
<i>N</i>	48	24	24	48	24	24	48	24	24			

Note: BAWL-R, Affective Word List Reloaded.

** $p < .01$.

Table 4. *Correlations of arousal ratings*

	Child Ratings			Laboratory			Adult Ratings Online Survey			BAWL		
	All Words	Positive Words	Negative Words	All Words	Positive Words	Negative Words	All Words	Positive Words	Negative Words	All Words	Positive Words	Negative Words
Child r_s Ratings				.676**	.512*	.745**	.538**	.266	.651**	.571**	.433*	.550**
<i>p</i>				.000	.011	.000	.000	.210	.001	.000	.035	.005
<i>N</i>				48	24	24	48	24	24	48	24	24
Adult r_s Ratings lab	.676**	.512*	.745**				.749**	.600**	.760**	.704**	.537**	.825**
<i>p</i>	.000	.011	.000				.000	.002	.000	.000	.007	.000
<i>N</i>	48	24	24				48	24	24	48	24	24
Adult r_s Ratings online survey				.749**	.600**	.760**				.502**	.256	.663**
<i>p</i>				.000	.002	.000				.000	.227	.000
<i>N</i>				48	24	24				48	24	24
Adult r_s Ratings BAWL	.571**	.433*	.550**	.704**	.537**	.825**	.502**	.256	.663**			
<i>p</i>	.000	.035	.005	.000	.007	.000	.000	.227	.000			
<i>N</i>	48	24	24	48	24	24	48	24	24			

Note: BAWL-R, Affective Word List Reloaded.

* $p < .05$. ** $p < .01$.

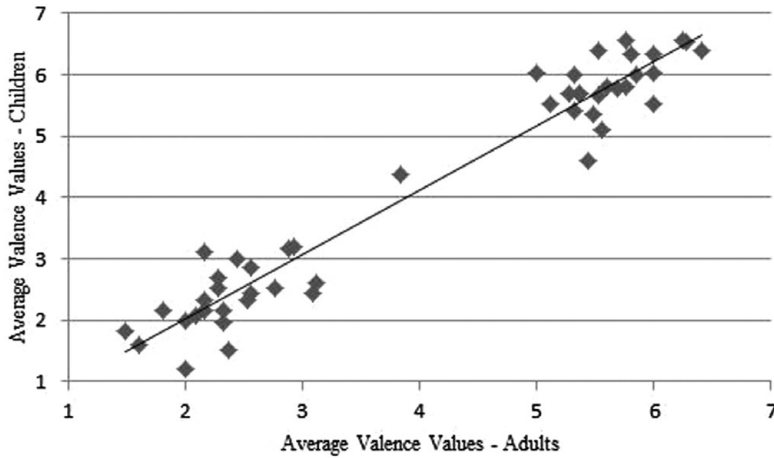


Figure 1. Children's ratings of valence as a function of the adults' norms.

RESULTS

Correlation of ratings across settings and age

Concerning valence, the results revealed very strong positive correlations between all adult groups (see Table 3). For arousal, we also found significant and moderately strong correlations (see Table 4). These highly significant correlations support the validity of the data and suggest that ratings of valence and arousal are stable despite different methods of data assessment: adult participants in the BAWL-R study had to rate a mixed set of 250–350 neutral, affective, or emotion words, whereas adults in our study rated 48 emotion terms exclusively.

To answer the question of whether 9-year-old children rate valence and arousal of emotion terms similarly to adults, the correlations between the children and the adult laboratory group are most interesting. As expected, the ratings for valence showed a very strong positive correlation over all items ($r_s = .91, p < .001$; see Figure 1 and Table 3), indicating that the adult ratings strongly predict those of the children. Even with a more detailed look at the subgroups of positive and negative words, the correlations turned out to be moderately strong for positive ($r_s = .55, p < .01$) and for negative ($r_s = .69, p < .001$) words. In addition, we calculated split-half reliabilities for valence ratings in children and adults. Results indicated a high agreement of raters within each age group (children: $r_s = .95, p < .000$; adults: $r_s = .94, p < .000$). Judgments regarding arousal showed moderately strong significant correlations ($r_s = .68, p < .001$; positive words only, $r_s = .51, p < .02$; negative words only $r_s = .75, p < .001$; see Figure 2 and Table 4). Split-half reliability was moderately strong for children's ratings ($r_s = .45, p < .001$) and strong in adults' ratings ($r_s = .92, p < .000$).

Comparison of absolute values

Following this, we compared the absolute rating values between the age groups in order to further investigate similarities and differences in the way 9-year-olds and

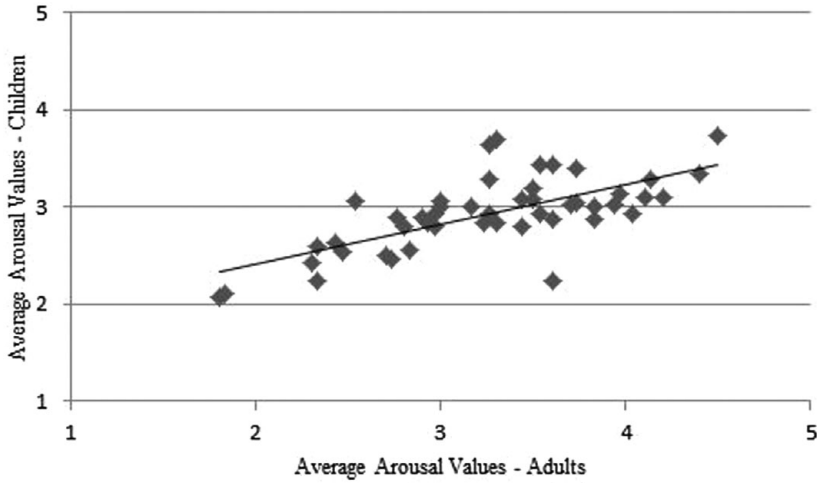


Figure 2. Children's ratings of arousal as a function of the adults' norms.

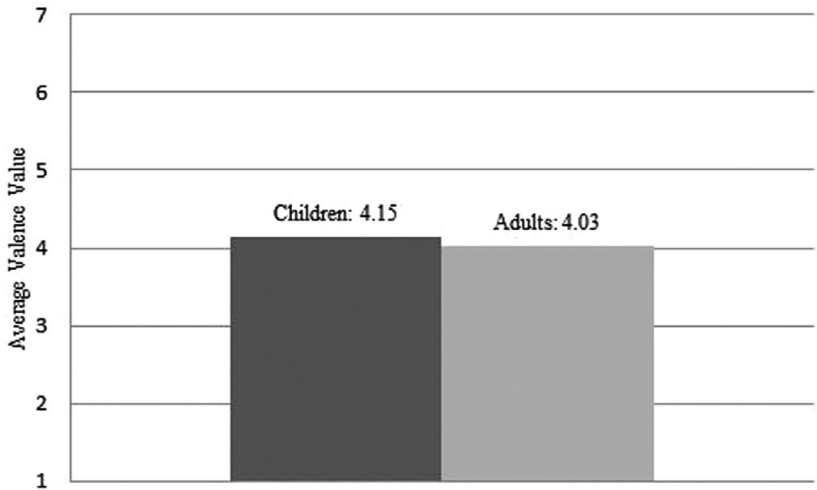


Figure 3. Average valence rating values of children and adults.

adults perceive the valence and arousal of emotion terms. For the analysis, we only considered the rating data of children and adults that were collected in laboratory settings.

Regarding valence, children tended to rate the words more positively than adults (see Figure 3). However, no main significant effect of age was found between the two groups: $F(1, 58.34) = 3.24, p = .077, d = 0.06$. In addition, a significant interaction of age and valence (positive or negative words) was observed,

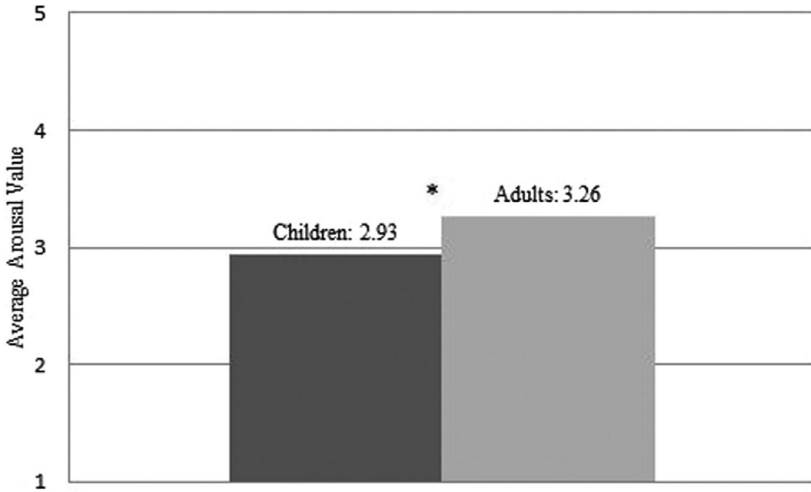


Figure 4. Average arousal rating values of children and adults.

$F(1, 2739.46) = 5.58, p = .000$, reflecting that the slight but insignificant difference between children and adults was more pronounced in the case of positive words (children: $M = 5.87$, of a scale of 1 to 7, $SD = 0.41$; adults: $M = 5.67, SD = 0.71$) than negative words (children: $M = 2.37$, of a scale of 1 to 7; adults: $SD = 0.44, M = 2.39, SD = 0.44$). Post hoc comparisons (t tests for unpaired samples) for the subsets of positive and negative words (24 words in each subset) did not produce significant results: positive words, $t(58) = 1.30, p = .200$; negative words, $t(58) = 0.10, p = .921$.

With respect to the mean arousal rating values, a significant main effect of age appeared, $F(1, 57.96) = 5.24, p = .026, d = 0.61$ (see Figure 4): children rated the word set as less arousing than adults.

Taken together, the analysis of children's and adults' ratings of 48 German emotion words with respect to valence and arousal revealed the following main results. Strong positive correlations between children and adults were observed for valence ratings as well as arousal ratings. With respect to absolute values, findings for valence showed that children and adults did not differ in how positively or negatively they perceive the meaning of emotion terms. In contrast, adults tended to rate words as slightly more arousing than children.

DISCUSSION

The goal of the present study was to collect children's valence and arousal norms for a set of 48 positive and negative German emotion terms controlled for several linguistic variables. To start with, adults' norms for valence and arousal offered by the BAWL-R and the two types of adult ratings collected in the present study were significantly correlated. This result demonstrates that ratings for valence

and arousal are replicable with distinct samples of adult participants and across different settings (controlled situation in the lab or online survey).

Having established the validity of the procedure, our main focus was to investigate whether 9-year-olds already perceive both emotional dimensions similarly to adults. In order to address this question, we not only assessed correlations of children's and adults' ratings but also carried out an analysis of absolute values of both dimensions for the two age groups.

For the valence ratings, the results demonstrate a very strong and positive correlation between children and adults, which was still noticeable when positive and negative words were analyzed separately. Arousal ratings of emotion words were also positively correlated between adults and 9-year-olds, although this correlation was only moderately strong. Our findings are in line with the results of Sylvester et al. (2016) as well as with those of Russell and Paris (1994), who found significant correlations between adults and children (age 7 to 12, respectively, 4 and 5 years). High correlations between children's and adults' ratings of valence and arousal have also been shown in rating studies using the same SAM scales with emotional stimuli from other modalities (e.g., McManis, Bradley, Berg, Cuthbert, & Lang, 2001, for pictures; Vesker, Bahn, Degé, Kauschke, & Schwarzer, 2017, for faces). In addition, our findings confirm the results of Sylvester et al. (2016), Schmidtke et al. (2014), and Russell and Paris (1994) with respect to the differential strength of the correlation of arousal and valence ratings. As expected, the adult-child correlation obtained in the arousal rating was weaker than in the valence rating. This finding might partly be because split-half reliability for arousal was much lower than for valence in the children's group (a similar finding was reported by Warriner, Kuperman, & Brysbaert, 2013). Given that internal reliability for arousal was already low, correlations between different sets of arousal judgments may not be expected.

The correlational findings strongly suggest that children were sufficiently familiar with the emotion terms and that they seem to organize them in much the same way as adults (as already stated by Russell & Ridgeway, 1983, for emotional grouping tasks). Since children and adults rated the valence and arousal of emotion words in the same direction on the two bipolar scales, they likely followed similar perception patterns.

In examining absolute values for valence and arousal from children and adults, a somewhat different picture emerged. Although no significant differences between absolute values of valence obtained from children and adults could be found, children's ratings showed a tendency to be more positive (especially for positive words). A related study in which we collected adults' and children's ratings of emotional facial expressions (Vesker et al., 2017) led to similar results, suggesting that, regardless of valence and modality, children might tend to perceive emotionally toned stimuli more positively than adults.

Turning to absolute arousal rating values, our results showed that children rated words as less arousing than adults. This finding is comparable to the reported rating studies with young, middle-aged, and older adults (Gilet et al., 2012; Grunwald et al., 1999; Söderholm et al., 2013) in which younger participants rated words lower on arousal than older participants. However, it runs contrary to Sylvester et al. (2016), who found the opposite: higher ratings on arousal for the younger age group

(children) in comparison to adults. The contrasting results may be due to item-specific differences between the two studies. Whereas the children investigated by Sylvester et al. (2016) were asked to rate affective words (a mixture of words with concrete and abstract meanings, as well as emotion terms), the children in our study exclusively rated a set of emotion terms. Therefore, some of the word meanings in our study are probably more abstract and acquired later in the course of development. Thus, their semantic representation in the children's mental lexicon might be less differentiated. This lower degree of differentiation could be the reason for children perceiving those words as less arousing than adults. It should also be kept in mind that the numerical difference between children and adults with respect to arousal ratings, although being significant, was only marginal in our study. Therefore, caution is warranted when interpreting this finding.

Conclusion

The present study was aimed at investigating commonalities and differences in adults' and children's judgments of valence and arousal for German emotion terms. Taken together, children showed a nearly adultlike judgment behavior when estimating the valence and arousal of emotion terms. Certainly, emotion terms represent a rather small, but nevertheless important, part of the vocabulary. However, in conjunction with the findings from children's and adult's norms for affective words (obtained by Sylvester et al., 2016), our results strongly suggest that children at the age of 9 are already aware of emotional properties conveyed by words. In sum, the commonalities between children's and adults' ratings seem to predominate over the small age difference in the arousal ratings and the tendency for children to make more positive evaluations.

Overall, our results justify the usage of adults' norms for controlling emotional word stimuli in studies with 9-year-old and older children, especially when children's norms are lacking. Nevertheless, the use of child-based norms in studies with children is clearly preferable. Additional studies with younger participants are also needed, since the present data does not allow for conclusions concerning younger children's perception of emotion terms.

ACKNOWLEDGMENTS

Support for this research was provided by the German Research Foundation (DFG). This research is part of the DFG-funded Collaborative Research Center Cardinal Mechanisms of Perception (SFB/TRR135 540 135/1 2014) at the Universities of Marburg and Gießen, Germany. We thank all families and students for their participation; Franziska Degé for her support in the project; and Johanna Sommer, Isabell Debus, Cecilia Sweitzer, and Gianna Preussner for their help in data collection.

REFERENCES

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412. doi:[10.1016/j.jml.2007.12.005](https://doi.org/10.1016/j.jml.2007.12.005)

- Baayen, R. H., Gulikers, L., & Piepenbrock, R. (1995). The Celex lexical database (Release 2, German Version 2.5, English Version 2.5, Dutch Version 3.1). Philadelphia, PA: Linguistic Data Consortium.
- 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*, 25, 109. doi:[10.3389/fnevo.2010.00109](https://doi.org/10.3389/fnevo.2010.00109)
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25, 49–59. doi:[10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
- Bradley, M. M., & Lang, P. J. (1999). *Affective Norms for English Words (ANEW): Instruction manual and affective ratings* (Technical Report C-1). Gainesville, FL: University of Florida, Center for Research in Psychophysiology.
- Bretherton, I., & Beeghly, M. (1982). Talking about internal states: The acquisition of an explicit theory of mind. *Developmental Psychology*, 18, 906–921.
- Citron, F., Weekes, B. S., & Ferstl, E. C. (2014). How are affective word ratings related to lexicosemantic properties? Evidence from the Sussex Affective Word List. *Applied Psycholinguistics*, 35, 313–331. doi:[10.1017/S0142716412000409](https://doi.org/10.1017/S0142716412000409)
- Foolen, A. (2015). Word valence and its effects. In U. M. Lüdtke (Ed.), *Consciousness and emotion: Vol. 10. Emotion in language* (pp. 241–256). Amsterdam: John Benjamins.
- Gilet, A.-L., Grünh, D., Studer, J., & Labouvie-Vief, G. (2012). Valence, arousal, and imagery ratings for 835 French attributes by young, middle-aged, and older adults: The French Emotional Evaluation List (FEEL). *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, 62, 173–181. doi:[10.1016/j.erap.2012.03.003](https://doi.org/10.1016/j.erap.2012.03.003)
- Grunwald, I. S., Borod, J. C., Obler, L. K., Erhan, H. M., Pick, L. H., Welkowitz, J., . . . Whalen, J. (1999). The effects of age and gender on the perception of lexical emotion. *Applied Neuropsychology*, 6, 226–238. doi:[10.1207/s15324826an0604_5](https://doi.org/10.1207/s15324826an0604_5)
- Hofmann, M. J., Kuchinke, L., Tamm, S., Vö, M. L. H., & Jacobs, A. M. (2009). Affective processing within 1/10th of a second: High arousal is necessary for early facilitative processing of negative but not positive words. *Cognitive, Affective, & Behavioral Neuroscience*, 9, 389–397. doi:[10.3758/9.4.389](https://doi.org/10.3758/9.4.389)
- Imbir, K. K. (2016). Affective norms for 4900 Polish words reload (ANPW_R): Assessments for valence, arousal, dominance, origin, significance, concreteness, imageability, and age of acquisition. *Frontiers in Psychology*, 7, 1081. doi:[10.3389/fpsyg.2016.01081](https://doi.org/10.3389/fpsyg.2016.01081)
- Jacobs, A. M., Vö, M. L.-H., Briesemeister, B. B., Conrad, M., Hofmann, M. J., Kuchinke, L., . . . Braun, M. (2015). 10 years of BAWLing into affective and aesthetic processes in reading: What are the echoes? *Frontiers in Psychology*, 6, 714. doi:[10.3389/fpsyg.2015.00714](https://doi.org/10.3389/fpsyg.2015.00714)
- Janschewitz, K. (2008). Taboo, emotionally valenced, and emotionally neutral word norms. *Behavior Research Methods*, 40, 1065–1074. doi:[10.3758/BRM.40.4.1065](https://doi.org/10.3758/BRM.40.4.1065)
- Kanske, P., & Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Research*, 1148, 138–148.
- 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, 634–645. doi:[10.1002/icd.1767](https://doi.org/10.1002/icd.1767)
- Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General*, 143, 1065–1081. doi:[10.1037/a0035669](https://doi.org/10.1037/a0035669)
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44, 314–324. doi:[10.3758/s13428-011-0168-7](https://doi.org/10.3758/s13428-011-0168-7)
- McManis, M. H., Bradley, M. M., Berg, W. K., Cuthbert, B. N., & Lang, P. J. (2001). Emotional reactions in children: Verbal, physiological, and behavioral responses to affective pictures. *Psychophysiology*, 38, 222–231. doi:[10.1111/1469-8986.3820222](https://doi.org/10.1111/1469-8986.3820222)

- 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, 169–177. doi:10.3758/s13428-012-0243-8
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.
- Russell, J. A., & Bullock, M. (1986). On the dimensions preschoolers use to interpret facial expressions of emotion. *Developmental Psychology*, 22, 97–102. doi:10.1037/0012-1649.22.1.97
- Russell, J. A., & Paris, F. A. (1994). Do Children acquire concepts for complex emotions abruptly? *International Journal of Behavioral Development*, 17, 349–369.
- Russell, J. A., & Ridgeway, D. (1983). Dimensions underlying children's emotion concepts. *Developmental Psychology*, 19, 795–804.
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44, 695–729. doi:10.1177/0539018405058216
- Schmidtke, D. S., Schröder, T., Jacobs, A. M., & Conrad, M. (2014). ANGST: Affective norms for German sentiment terms, derived from the affective norms for English words. *Behavior Research Methods*, 46, 1108–1118. doi:10.3758/s13428-013-0426-y
- Schröder, S., Würzner, K.-M., Heister, J., Geyken, A., & Kliegl, R. (2015). childLex: A lexical database of German read by children. *Behavior Research Methods*, 47, 1085–1094. doi:10.3758/s13428-014-0528-1
- Schwarz-Friesel, M. (2013). *UTB Sprachwissenschaft: Vol. 2939. Sprache und emotion* (2nd ed.) Tübingen, Germany: Francke.
- Söderholm, C., Häyry, E., Laine, M., & Karrasch, M. (2013). Valence and arousal ratings for 420 Finnish nouns by age and gender. *PLOS ONE*, 8, e72859. doi:10.1371/journal.pone.0072859
- 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, 22. doi:10.3389/fpsyg.2016.00969
- Syssau, A., & Monnier, C. (2009). Children's emotional norms for 600 French words. *Behavior Research Methods*, 41, 213–219. doi:10.3758/BRM.41.1.213
- 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, 1169–1180. doi:10.1016/j.paid.2006.03.025
- Vesker, M., Bahn, D., Degé, F., Kauschke, C., & Schwarzer, G. (2017). Perceiving arousal and valence in facial expressions: Differences between children and adults. *European Journal of Developmental Psychology*. Advance online publication. doi:10.1080/17405629.2017.1287073
- Vigliocco, G., Kousta, S.-T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2013). The neural representation of abstract words: The role of emotion. *Cerebral Cortex*, 24, 1767–1777. doi:10.1093/cercor/bht025
- Võ, M. L. H., 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, 534–538. doi:10.3758/BRM.41.2.534
- Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45, 1191–1207. doi:10.3758/s13428-012-0314-x
- Widen, S. C., & Russell, J. A. (2010). Differentiation in preschooler's categories of emotion. *Emotion*, 10, 651–661. doi:10.1037/a0019005