Applied Psycholinguistics (2023), **44**, 29–50
doi:10.1017/S0142716422000431

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

Assessment of emotion word vocabulary and its contribution to reading comprehension

Tami Sabag-Shushan[1,2,3,*], Tami Katzir[1] and Anat Prior[1]

[1]Edmond J. Safra Brain Research Center, Department of Learning Disabilities, University of Haifa, Haifa, Israel, [2]Department of Education, Ben-Gurion University of the Negev, Beer Sheva, Israel and [3]Center for the Study of Pedagogy – Research Practice Partnerships, Ben-Gurion University of the Negev, Beer Sheva, Israel
*Corresponding author. Email: Tammisho@gmail.com

(Received 24 July 2021; revised 31 July 2022; accepted 4 November 2022; first published online 05 January 2023)

Abstract
The contribution of vocabulary to academic achievements in general and to reading comprehension (RC) in particular has led to the development of various tools for vocabulary assessment. However, existing assessments do not distinguish between word types, and specifically, they do not target emotion vocabulary, despite growing recognition of the importance of emotional processing to RC ability. In this study, we first describe the development of a novel vocabulary assessment in Hebrew – Herut – and examine its validity and sensitivity. This assessment includes both emotion words and nonemotion words, and is based on curriculum. Next, we studied the contribution of the emotion and nonemotion words subscales of the Herut to RC in a sample of 1,333 Hebrew-speaking fourth- and fifth-grade students. Both types of vocabulary knowledge made significant independent contributions to RC, and the contribution of emotion words was slightly larger than that of nonemotion words. Finally, the Herut measure was found to be more predictive of RC than a general vocabulary measure in Hebrew.

Keywords: vocabulary; emotion words; curriculum-based assessment; reading comprehension

Introduction

The considerable contribution of vocabulary knowledge to reading comprehension (RC) (Pearson et al., 2020) has led to the development of receptive and expressive vocabulary assessments in different languages. Research reviewing extant vocabulary assessment tasks has emphasized the importance of defining the purpose of the test, describing the development procedures both theoretically and practically, as well as the need for stricter validation procedures (Pearson et al., 2007; Schmitt et al., 2020). The importance of vocabulary knowledge is also evident by the fact that academic word lists, which set out explicit instruction goals in this domain, have been compiled for different grade levels (Coxhead, 2000, 2011; Gardner & Davies, 2013). However, existing assessment measures and word lists typically

approach vocabulary as a single construct without distinguishing between word types, and in particular, do not include emotion words (Coxhead, 2000, 2011). The distinctions between emotion and nonemotion words are supported by broad theoretical justifications. Emotion words are different from both other abstract and concrete words in their tangibility, imageability, and degree of connection and availability in a semantic network (Altarriba & Basnight-Brown, 2010; Altarriba et al., 1999; Pavlenko, 2008). Differences between emotion words and nonemotion words are also evident in information processes; for example, emotion words were found to be better recalled than nonemotion words (Kissler et al., 2007; Kissler et al., 2009).

From a developmental perspective, emotion words are acquired at later ages, usually after entering school, compared to general verbs (for example, action verbs or movement), and even later than the development of cognitive words (e.g., "thinking" and "conclusion") (Ravid & Egoz-Liebstein, 2012). Given these differences, it is important to study the role of emotion words in RC as well, since it may diverge from that of other types of words. In the present study, we developed a curriculum-based theoretically motivated vocabulary assessment in Hebrew, with a specific focus on emotion words. We then tested how well this assessment might predict RC compared to a previously validated measure of general vocabulary.

Vocabulary and RC

Vocabulary is the conceptual knowledge of words and their meanings (Argaman, 2006; Perfetti, 2007; Silverman et al., 2014). Readers need to understand the meaning of the words they encounter in texts, in order to be able to construct their meaning. This is especially important in higher grades, since text complexity increases and requires knowledge of more sophisticated vocabulary (Neuman & Wright, 2014). A well-established base of broad and varied vocabulary contributes to the comprehension of unfamiliar words as well (Ricketts et al., 2011; Shahar-Yames & Prior, 2018). A meta-analysis conducted by the US Institute of Education Sciences (Pearson et al., 2020) found that vocabulary was the most powerful predictor of RC. However, one of the critical conclusions of the study was that while the different vocabulary measures included across studies were robust in predicting RC, there has not been a systematic effort in the reading field to justify the procedures of selecting words for assessment and instruction. Others have also expressed the critique that many vocabulary assessment measures are not theory-based (Pearson et al., 2007) and refer to vocabulary as a single general structure without addressing different types of vocabulary.

Vocabulary assessment

The contribution of vocabulary to academic achievements in general and to RC in particular has led to the development of various tools for assessing this knowledge, as well as to the development of academic word lists. These lists contain words that are essential for comprehension in various disciplines (Coxhead, 2000) and are based on the frequency of occurrence in academic materials and content databases (Coxhead, 2000, 2011; Gardner & Davies, 2013). One of the best-known lists is the Academic Words List, published in 2000 by Coxhead for university students.

https://doi.org/10.1017/S0142716422000431 Published online by Cambridge University Press

Subsequent word lists were also developed for younger students, such as the Oxford wordlist used in Australian primary schools (2021). Most of these lists refer to vocabulary as a single general construct and do not identify different types of words.

Similarly, the educational field of vocabulary assessment has yet to offer distinctions between vocabulary types. Over the years, principles for word choice were formulated for assessment tasks. The National Assessment Governing Board (NAGB, 2005) defined a set of word selection criteria: words which are the characteristics of mature language and appear in written language; words that are not basic or common but also not uncommon, or that reflect the specific subject (Beck et al., 2013); words that are found necessary for understanding the main ideas in texts; and words found in learners' reading material. However, there are studies that indicate gaps and limitations both in the measurement procedures and in how the tools are validated (Schmitt et al., 2020), as well as in how words are chosen for common vocabulary assessment tools (Pearson et al., 2007).

The first gap is related to methodological aspects. Schmitt et al. (2020) claim that background materials published with vocabulary measures provide no theory or guidance on how words were chosen. Thus, there is no way of knowing why these specific items were chosen rather than others (Pearson et al., 2007) and whether assessment measures should be based on vocabulary from the curriculum or on another criterion, such as the frequency of the word in the language.

The second gap is related to validation. Schmitt et al. (2020) evaluated vocabulary measures and found that most measures were launched with poor or insufficient evidence as to their validity. Thus, it is important to develop vocabulary assessment tools while adhering to higher requirements and standards of tool validity.

Beyond the gaps in vocabulary measure development procedures and how they are validated, extant vocabulary measures also do not include different types of words. The importance of integrating diverse words into vocabulary assessment is also expressed in the National Assessment of Educational Progress (NAEP, 2009), which emphasized that selected words should be necessary for a variety of contexts and areas and can express concepts, ideas, actions, or emotions.

It seems that even in widely used vocabulary measures, such as the Wechsler Intelligence Scale for Children (WISC) (Kaufman, 1994) and the Peabody Picture Vocabulary Test (PPVT) (Dunn & Dunn, 1981), insufficient information is given on how words were selected to be included in the test. The PPVT Manual (Dunn, 1965) for example does not include a comprehensive explanation of how the words are chosen. Thus, it is noted that out of a collection of 3,885 words taken from a dictionary, 2,055 words from a variety of topics were utilized, but there is no explanation as to why specific words were chosen over others. In a more recent version of the PPVT (PPVT-4_2007), there is only general information about the choice of words for the test. For example, it is stated that the words cover a wide range of vocabulary levels. The words represent 20 content areas (e.g., actions, vegetables, tools) and parts of speech. However, there is no detailed information about how the words were chosen in each category or information about word frequency and length. Similar gaps can be identified in common vocabulary tests in Hebrew and in additional languages. One common vocabulary measure in Hebrew comes from the Elul battery (Shatil et al., 2007). In this test, a word is presented in each question and the students are asked to identify the picture that represents the word

https://doi.org/10.1017/S0142716422000431 Published online by Cambridge University Press

most appropriately. According to the description, in order to select the words, the research team was exposed to diverse items in the professional literature. However, it is not clear exactly what professional literature was at the base of the selection process and how these items were selected. The manual also mentions that in formulating the items for the test, each item was run through a sequence of tests, but it is unclear what these tests entailed and what were the criteria for selecting certain items over others. The Elul battery was administered to students from 120 classes from different schools. Based on the data collected, norms for each grade level were calculated for each subtest.

In a recent revision of the Elul test, conducted by Shany and colleagues (2017), all the words appearing in the original subtests for the third and sixth grades versions of the test were administered to 20 students from each age grade level (from first to sixth grade). Based on the accuracy of the various age groups, the original items were ranked by level of difficulty. The revised test includes a single list of words to be administered to all age groups, and words were presented with increasing level of difficulty. The test begins with common words in the language (such as "jump"), and towards the end of the test, more complex, specific, and rarer Hebrew words are presented (such as "cycle"). However, the revised version relies on the words included in the original Elul subtest, again, for which there is only scant information about how they were identified and selected. To the best of our knowledge, emotion words have not yet been found systematically in vocabulary measures.

Characterizing different emotions

According to Ekman's model of emotions (1972), basic emotion develops in early childhood and often arises as a result of direct and specific circumstances that are easy to identify from facial expressions, even without knowing the context (Ekman, 1982; Lewis, 1993a). Ekman identified six basic emotions (sadness, fear, disgust, joy, anger, and surprise). In contrast, nonbasic emotions arise as a result of more complex circumstances and require sophisticated cognitive ability to understand (Lewis, 1993b). They develop at a later stage and continue to develop during the school-age years (Harris, 1993; Lewis, 1993a). The assessment task developed for this study included both basic and nonbasic emotions

According to Scherer's (2005) Alternative Dimensional Structures of the Semantic Space for Emotions model, emotions can be mapped across four dimensions: valence (positive vs. negative), level of arousal (active vs. passive), degree of activation (obstructive vs. conducive), and degree of control (high vs. low). These four dimensions form a wheel divided into eight parts. Each part reflects two main characteristics and includes several emotions within it. One slice includes all the emotions that correspond to the active qualities with high control (e.g., lust, pride, ambition). The second, opposite slice, includes the set of emotions that correspond to passive characteristics and low control (for example, indifferent, hesitant, embarrassed, bored, and sad). The third slice includes the set of emotions that correspond to traits of negative values and low control (e.g., lonely, despairing, depressed, dissatisfied, disappointed, and unhappy). The opposite fourth slice includes emotions with positive values and controllability (e.g., enthusiastic, eager, interested, amused, and happy). The next slice contains negative values and a degree of

https://doi.org/10.1017/S0142716422000431 Published online by Cambridge University Press

activation (e.g., frustrated, suspicious, rejected, anxious). Opposite to this are positive emotions and high degree of activation (e.g., in love, confident, hopeful, happy, calm, sociable, satisfied). The final two slices relate to a high degree of arousal (active) with low degree of activation (e.g., jealous, afraid, angry) and opposite it a low degree of arousal (passive, calm) with a high degree of activation (e.g., longing, empathic). In the current study, we adopted Scherer's model (2005) for designing a vocabulary assessment tool in Hebrew, such that it included words from all eight slices of the model.

From emotions to emotion words

Emotion words are recognized as a lexical choice that directly reflects an emotional state, for example, joy, pride, and frustration (Pavlenko, 2008). Emotion words provide information about inner states, attitudes, beliefs, social connections, motives, values, and behaviors (Shiota & Kelther, 2005). The words can be negative or positive and can be connected to the arousal of physiological responses (Pavlenko, 2008; Hsu et al., 2015). This definition does not include emotionally laden words which describe expressions of emotion indirectly, such as curses or affectionate phrases (Pavlenko, 2008). Emotion words express the connection between the subjective experience of an individual and her/his environment on the one hand and the emotional experience and her/his future action tendencies on the other (Hobbs & Gordon, 2011). Emotion words are often very meaningful and play an important role in the perception of experiences and interpretation of the individual in diverse emotional states (Gendron et al., 2012).

The ability to name emotion vocabulary is a basic ability among a variety of emotional abilities (Denham, 2007; Harris, 1993; Izard et al., 1996). Emotion words are most often explored in the context of social behavior (Izard et al., 2001; Miller et al., 2005), and they may reflect cultural differences in the emotional lexicon (Fontaine et al., 2013).

Emotion words vs. nonemotion words

Emotion vocabulary has been less explored in the context of cognitive and reading processes and rarely in the context of vocabulary assessment. We hypothesize that there is a strong basis for treating emotion vocabulary as separate from nonemotion vocabulary for several reasons.

The first reason is that emotion words have unique characteristics. Emotion words are conceived as mental words, namely abstract words that describe internal processes or situations, including emotions (Hall & Nagy, 1986; Ravid & Egoz-Liebstein, 2012; Ruffman et al., 2002; Shiota & Keltner, 2005). Mental words have abstract and complex meanings based on the comprehension of mental states and are challenging to learn from context (Ravid & Egoz-Liebstein, 2012). People report more difficulty in representing emotion words in their imagination compared to abstract words. Emotion words are also more challenging to recall than concrete words or abstract words (Altarriba & Basnight-Brown, 2010; Altarriba et al., 1999; Pavlenko, 2008). In addition, written emotion words were detected very rapidly and automatically about 250 ms after exposure, unlike neutral words that took 500 ms (Kissler et al., 2007).

https://doi.org/10.1017/S0142716422000431 Published online by Cambridge University Press

The second reason to include emotion words in vocabulary assessment is research that suggests that they have a different course of language acquisition rate compared to nonemotion words. Mental verbs have been consistently acquired relatively late compared to nonmental verbs (Ravid & Egoz-Liebstein, 2012). Within mental words, it seems that emotion words are acquired especially late. Studies have found that children aged 5–8 were more likely to use cognitive verbs (such as thinking and knowing) than emotion verbs (such as loving and feeling sad). The later acquisition of emotion words may be related to the later development of abstraction abilities (Egoz-Liebstein, 2009; Ravid & Egoz-Liebstein, 2012). We could not identify studies examining the proficiency of older children with emotion words.

These differences reinforce the notion that emotion words might function independently than other types of words and provide the theoretical basis for assessing emotion words as a specific category of vocabulary and examining their contribution to RC.

The current study

In the current study, we developed a novel vocabulary assessment (the Herut test) that attempts to address the gaps identified in the literature. The assessment we developed is curriculum-based and includes different categories of emotions according to Scherer's model (2005).

The main goal of the current study was to validate this newly developed vocabulary measure (Herut) and examine the contribution of emotion words to RC. To achieve an equal balance of 3 emotions from each of the 8 slices in the Scherer's model, 24 emotion words were selected from the curriculum. According to Ekman's model (1972), another 6 basic emotions were added, for a total of 30 emotion words. These items were split between two versions of the test, one of which was used in the present study. Because the current study was focused specifically on emotion words in the context of RC abilities and based on the theoretical framework presented above that reflects different types of emotion words, the Herut measure we designed included more emotion than nonemotion words.

After developing the measure, we next examined its psychometric characteristics. The first aim will be to validate the Herut measure. To examine *convergent validity*, Pearson correlations were calculated between the Herut measure and the Elul vocabulary measure. Elul and Herut are both receptive measures of vocabulary, but Herut is curriculum-based, whereas Elul is based on increasing level of difficulty and is age normed. In addition, Herut requires children to match words to definitions, whereas Elul requires matching words to visual representations. Based on the research conducted by Beck et al. (2012), which found a correlation between emotional abilities and linguistic abilities such as vocabulary, we hypothesized that there would be positive correlations between the two vocabulary measures. We also examined the correlation between the two subscales within the Herut measure, namely emotion and nonemotion words.

In the next step, we assessed the construct validity of Herut. To this end, we calculated Pearson correlations between the Herut and reading fluency and RC measures. Based on Beck et al. (2012), we hypothesized that there would be positive

https://doi.org/10.1017/S0142716422000431 Published online by Cambridge University Press

correlations between the two subscales of the Herut measure (emotion and nonemotion words) and reading fluency and RC measures.

The second question addressed in the current study is the contribution of vocabulary measures and words types (emotion and nonemotion words) to predicting RC by using a hierarchical regression analysis. To our knowledge, no studies have evaluated the differences between the contributions of different vocabulary measures to RC, so there was no basis for an apriori hypothesis.

Development of the Herut vocabulary measure

The Herut test is a unique tool aimed to assess different types of vocabulary in Hebrew-speaking upper elementary school aged children (grades 4 and 5). The purpose of the assessment is to evaluate knowledge of emotion and nonemotion words, based on the curriculum.

The measure was developed in several steps: As a first step, texts were selected from fourth-grade curriculum materials, including textbooks, assessment tasks, and teaching units published by the Ministry of Education and independent publishers approved by the Ministry. The selected texts were either narrative or expository and were written by various authors. The sample included texts from different domains: RC and literature (15,000 words), science (8,000 words), geography (5,000 words), and social skills (5,000 words). All texts together included 33,441 words (about half of them narrative and half expository). Based on accepted pedagogical principles (Beck et al., 2013; Graves et al., 2013), candidate words for the vocabulary assessment were identified first by the first author and then by two trained research assistants. We selected second-tier words (such as "embarrassment," "inspiration"), which are not basic or highly frequent words on the one hand, but are also not rare words specialized for a specific topic (Beck et al., 2013). Of the selected candidate words, some appeared explicitly in the text, and some were imported words (words that did not appear in the actual text but were deemed to be important for supporting understanding of the text, Graves et al., 2013). For example, in a text that describes several people working together for a long time to fix something, the words "collaboration" and "goal" are necessary for understanding the main idea in the text even though the words themselves did not appear explicitly.

Based on these characteristics, 2,057 words (6%) were selected from the text corpus. The chosen words were categorized into two groups:

Emotion words: Words that describe emotions directly (e.g., expectation, loneliness, stress). (Pavlenko, 2008).

Nonemotion words: This category includes all other words that do not describe emotions. These words could be general words (e.g., collaboration, feature, advantage) and could also be cognitive words which reflect thought processes but not emotions (e.g., curiosity, definition, idea). Our choice in this classification stemmed from our specific focus on emotion words. Thus, words categorized as nonemotion words did not convey emotional information, but were of neutral, general content.

From the initially identified words, 397 words (or 19%) appeared in more than one text. Of these, 85% were nonemotion words, and 15% were emotion words. These 397 words were then rated by 40 regional language instructors who are

https://doi.org/10.1017/S0142716422000431 Published online by Cambridge University Press

employed by the Ministry of Education, and who supervise and provide training for elementary school language arts teachers. The regional language instructors were asked to rank the words' importance for teaching upper elementary school students, on a scale of *1 (not at all)* to *5 (very much)*. For low-ranked words, instructors were asked: Is the word too simple or complex? Is the word uncommon or not important? The questionnaire had three versions, each including about 130 words. In total, each version of the questionnaire was rated by 13 language instructors.

Following these steps, each word was ranked using six parameters: explicit frequency (number of appearances), frequency of words that did not appear in the actual text but were deemed to be important for supporting understanding of the text (imported words), the number of resource types in which the words appeared (curriculum materials, assessment tasks, teaching units), the number of content subjects in which the word appeared (language arts, literature, science, geography, and social skills), the number of text types (narrative/expository) in which the word appeared, and the word's average ranking by the language instructors. The average of each parameter was calculated across all words, and then each word that was above average received 1 point for the relevant parameter. Next, the scores were summed across these parameters for all words. The maximum score was 6 and the minimum was 0, with a mean of 2.72. Finally, the 100 words with the highest score were selected (scores of 4–6), 85 nonemotion words and 15 emotion words.

In the final stage, the selected emotion words were examined to determine whether they express the eight slices of emotion words based on the Semantic Space for Emotions model (Scherer, 2005). The emotion words identified in the previous steps were indeed represented in the model, but the list did not include equal representation for all eight slices. Therefore, additional words from the model were added to the test list. To do this, we identified six additional emotion words from the corpus-based list, which were not included following the initial selection process described above. These words received an average rating of 3 points. Finally, we added three additional words to fully reflect the categories of Scherer's (2005) model, which were not included in the corpus-based list. We selected words that were deemed appropriate for the target population in terms of content and frequency (e.g., words such as conscientious and gloomy were not included). In the final step, six basic emotion words were added based on Ekman's model (1972) (sadness, fear, disgust, joy, anger, and surprise) to avoid an overly complex level of difficulty. Thus, in order to remain faithful to Scherer's theory (Scherer, 2005) and to avoid a high level of difficulty, the final test included more emotion words than nonemotion words. The nonemotion words included in Herut were also selected from the list described above. We made sure to choose both words that reflect a cognitive state, such as the words comparison and opinion, and words that are not cognitive, such as the words advantage and sharing. A two-step process was taken in order to select the words. First words were rated by three independent judges and 13 language experts on several categories listed below. Then four independent judges choose words there were high in frequency on the listed parameters. The parameters were as follows:

The word appeared in a higher average frequency at the explicit and imported level (7.28 vs. 5.58).

https://doi.org/10.1017/S0142716422000431 Published online by Cambridge University Press

Table 1. Frequency of words in the adult corpus and in study materials

	Average frequency per million of adult corpus	Average of frequency in study materials	Average number of letters
Herut test – emotion words	175.36	4.55	4.1
Herut test – nonemotion words	174.22	6.22	5

The word appeared in a higher average of content subjects (2.76 vs. 2.62). The word appeared in a higher average of resource types (curriculum materials, assessment tasks, teaching units) (3.09 vs. 2.87) and a higher average ranking by the language instructors (4.49 vs. 4.36). No difference was found in the average of text types (narrative/expository) since all the selected words appeared in both types of texts.

In order to neutralize the effect of the frequency of words in language on the findings of the analysis, we examined the frequency of the word types: emotion words versus nonemotion words. First, the frequency of words from students' study materials in different disciplines was examined, as well as the frequency of words from a broader corpus, based on a database for word frequency in Hebrew (Frost & Plaut, 2005). Emotion and nonemotion words were well matched on frequency in the language and in the text corpus collected for this study, as well as on length in letters (see Table 1). In the *t* test for independent samples, no significant differences were found between the frequency of nonemotion words and the emotion words ($p > .05$).

From the final word list, two parallel versions (with different words) of the Herut measure were developed, but in the present study, data from only one of the Herut versions are analyzed. Each version contains 15 emotion words and 10 nonemotion words.

The Herut assessment had two parts. In the first part, students were presented with a target word and were asked to identify the correct definition from among four options. In the second part, students were presented with a definition and were asked to identify the correct word from among four options. The test was scored on accuracy – one point was given for each correct response. An accuracy percentage measure was calculated based on the number of correct items. Each student received a score for the entire task, and in addition a score for each scale separately – emotion, and nonemotion words.

In a sample of 1,333 Hebrew-speaking students, Cronbach's alpha for the entire measure (23 items) was .809, after two (nonemotion) items were removed due to low reliability. Cronbach's alpha of the emotion words (15 items) was .732 and of the nonemotion words (8 items) was .623.

Method

Participants

The sample in the present study was taken from a broad study (Katzir et al., 2019), which was conducted over 3 years, with a sample of 114 teachers and 2,352 students

https://doi.org/10.1017/S0142716422000431 Published online by Cambridge University Press

from 33 schools. School administrators were sent explanatory letters about the study, and only schools that expressed consent participated in the study. Letters of consent approved by Office of the Chief Scientist in the Ministry of Education were sent to students' parents. Only students whose parents signed a consent form, and who themselves gave consent, participated in data collection. Data were collected in school by trained research assistants who were studying for advanced degrees in education.

The sample in the present study was derived from the third year of the wider study. The research participants were 1,333 Hebrew-speaking students from 21 elementary schools (668 males and 665 females), 844 in fourth grade (63%), and 489 in fifth grade (37%). Eighty-five percent of participants were born in Israel. Participants varied in socioeconomic and sociodemographic characteristics: 236 students (18%) attended schools classified by the Ministry of Education as being of high socioeconomic status, 805 students (60%) attended schools classified as being of medium socioeconomic status, and 292 students (22%) attended schools classified as being of low socioeconomic status.

Measures

All the tools can be found at this link: https://osf.io/2XTRP/

Vocabulary

The Herut measure (Sabag-Shushan, Yosefi, Katzir & Prior). Described above.

The Elul Task (Shany, Prior & Blicher, 2017). A Hebrew vocabulary measure is based on a subset of items from the Elul Measure (Shatil et al., 2007). The task includes 35 words organized by frequency (high to low) in the Hebrew language and does not include emotion or cognitive words. For each target word, participants were required to choose the picture that most accurately matches the meaning of the target word from among four options. An accuracy percentage measure was calculated based on the number of correct items. Cronbach's alpha was .764.

Reading comprehension

Literal and inferential meaning (Shany & Blicher, 2017). Students read two texts, one narrative and one expository (416 words on average per text). After reading, students answered comprehension questions on two levels: 11 multiple choice questions that ask for explicit information in the text (simple comprehension) and 7 multiple choice questions examining inferences and integration with out-of-text information, including the student's prior knowledge or information from another text (complex comprehension). The narrative text included 10 questions ($\alpha = .717$) and the expository text 8 questions ($\alpha = .631$). Cronbach's alpha for the entire task (18 items) was .793.

Tamar Measure (Sabag-Shushan & Katzir, 2018). This is a RC measure to assess simple and complex comprehension abilities. Students read four short texts (two narrative and two expository, 145 words on average per text). Following each text, students answered five multiple choice questions: one question that evaluated

https://doi.org/10.1017/S0142716422000431 Published online by Cambridge University Press

explicit information (simple comprehension), three questions that assessed different types of mental inferences (inferences about the emotions of the characters in the text, the ability to evaluate the character's thoughts, and finally about the character's goal planning) and one question that evaluates the ability to identify a main idea in the text. The questions that assess inferences and main ideas are classified as complex comprehension. An accuracy percentage measure was calculated based on the number of correct items.

In order to illustrate the distinction between the types of questions, we present an example narrative text:

"A woman was shopping at the supermarket and because the weather was hot, she put her jacket in the shopping cart. After paying for the products, she loaded them into her car and then discovered that some additional products were hidden under the jacket. The woman quickly returned to the supermarket to pay for the missing products. The cashier told her: "Now I understand why the women behind you were whispering". The woman asked: "Where are these women?" The cashier answered that they had just left for the parking lot. The woman quickly rushed to the parking lot."

To assess emotional inference, we asked: What do you think the woman felt when she discovered the women were whispering about her? To assess the character's thoughts, we asked what the cashier understood when the woman returned to the supermarket. And to assess goal and planning inference, we asked why the woman rushed to the parking lot. Or in other words, this question evaluates the purpose of this behavior.

Similarly, in expository texts, we applied mental inference questions. For example, one text described the life of Alfred Nobel and his invention of dynamite, which was originally invented for the purpose of blasting rocks but was soon used for military purposes. The Nobel Prize Foundation was his way of encouraging people's contribution to humanity and to atone for the damage his invention led to. After reading the text, the students were asked what Alfred Nobel thought of the idea of using his invention for military purposes (inference about the character's thoughts), Why Alfred Nobel founded the Prize Fund (goal and planning inference), How did Alfred Nobel feel about his idea of founding the Nobel Prize Fund (emotional inference)?

Cronbach's alpha for the entire measure (20 items) was .833. Cronbach's alpha for the narrative texts (10 items) was .738, and for the expository texts, it was .689. The emotion and nonemotion words examined in the Herut vocabulary measure did not appear in the texts so there was no overlap between the words in the vocabulary measure and in the RC measure. Since this is a novel RC measure, and it evaluates, among other things, emotional inferences, we examined its correlations with the previously established Literal and Inferential meaning test (Shany & Blicher, 2017) and found them to be positively correlated ($r = .695$, $p < .001$).

Word reading fluency

The Hebrew version of the TOWRE reading fluency measure (Torgrsen et al., 1999), adapted by Katzir and her colleagues (2012), was administered individually to each student. This measure assesses decoding abilities that is, accuracy and reading

https://doi.org/10.1017/S0142716422000431 Published online by Cambridge University Press

Table 2. Descriptive statistics: range, mean, standard deviations, skewness, and kurtosis measures

	N	*Range*	*M*	*SD*	*Skewness*	*Kurtosis*
Reading fluency	1,204	0–103	55.37	16.79	−0.14	−0.32
Elul	1,236	0–100	67.22	14.19	−0.58	0.95
Herut	1,214	0–100	63.75	19.62	−0.54	−0.34
Emotion words	1,212	0–100	72.92	23.64	−0.94	0.12
Nonemotion words	1,214	0–100	56.86	24.98	−0.11	−0.84
RC – Tamar	1,216	0–100	54.00	24.00	−0.08	−1.12
RC – Literal and Inferential Meaning	1,290	0–100	56.74	23.77	−0.13	−0.97
RC – combined reading comprehension	1,292	2.63–97	53.58	23.79	−0.77	−1.08

Note. RC-simple and complex comprehension abilities, in narrative and expository text.

fluency. Participants are asked to read aloud a list of Hebrew words, presented in the nonpointed script, as fast as they could in 45 s. The words were ordered by increasing the level of difficulty. The score was the total number of words read accurately in 45 s. Cronbach's alpha was .769.

Procedure

Tasks were administered during two classroom sessions and one short session of individual testing. In each session, the students performed a RC task and a vocabulary task. In the individual session, the students completed the fluency task. All sessions were conducted within 3 weeks at most.

Results

Herut characteristics and performance

The first goal in this study was to validate the Herut test. We examined the normality of the distribution of scores, as well as Skewness and Kurtosis (Table 2).

Since more than one session was required, there were students who did not complete all the required assignments, so there are missing data. 1,333 reflects students had at least one data measure. 1,035 students completed the full battery.

Next, children's accuracy scores for the two types of vocabulary in the Herut test were compared. A paired-samples t test demonstrated that children were more accurate for emotion vocabulary ($M = 72.92$, $SD = 23.64$) than for nonemotion vocabulary ($M = 56.94$, $SD = 24.93$; $t_{(1,211)} = 25.66$, $p < .001$).

Validating the Herut test

Convergent validity

Expression of the convergent validity is through high correlations of two different measures (Herut and Elul), which measure a common feature, here vocabulary knowledge. The Herut measure and the Elul vocabulary measure were significantly

https://doi.org/10.1017/S0142716422000431 Published online by Cambridge University Press

Table 3. Correlations between vocabulary, fluency, and reading comprehension ($N = 1{,}204$)

	Reading fluency	RC – Tamar	RC – Literal and Inferential Meaning
Elul measure	.375*	.555*	.501*
Herut measure	.441*	.684*	.672*

*$p < .001$.

positively correlated ($r = .592$, $p < .001$). Positive correlations were also found between the Elul measure and each of the Herut subsets: emotion words ($r = .553$, $p < .001$) and nonemotion words ($r = .550$, $p < .001$). There was also a significant positive correlation between nonemotion words and emotion words within the Herut measure ($r = .602$, $p < .001$).

Construct validity

Expression of the construct validity is through correlation with related measures, such as reading fluency and RC indices. Pearson correlations between the Herut measure, reading fluency, and RC were calculated *N = 1,204)*. Significant positive correlations were found between all research variables (all $p < .001$) (Table 3).

Another expression of construct validity is through the assessment of differences between groups. Previous research has demonstrated that increasing age is associated with higher vocabulary knowledge (Seigneuric & Ehrlich, 2005). Thus, differences between fourth- and fifth-grade students in vocabulary knowledge on the Herut measure are another measure of test validity. An independent samples *t* test showed significant differences between grade levels in the Herut vocabulary measure ($t_{(1,212)} = -7.49$, $p < .01$), such that fifth graders showed higher accuracy ($M = 69.18$, $SD = 18.39$) than fourth graders ($M = 60.61$, $SD = 19.63$).

The contribution of different vocabulary measures to RC

Since the two RC tests (the Tamar test and the Literal and inferential meaning test) assess common aspects of RC ability, and because there is a high correlation between the tests (.695**), we calculated a combined RC measure based on both tests. In the combined measure, students read three texts of each type (narrative and informative) and answered 15 questions from the simple understanding ($\alpha = .788$) as well as 23 questions from the complex understanding version ($\alpha = .825$). Cronbach's alpha for the narrative tasks (20 items) was .831, and .789 for the informative tasks (18 items). Cronbach's alpha for the combined comprehension measure (38 items) was .894.

In order to examine the contribution of the Herut measure to predicting RC (the combined RC measure), a hierarchical regression analysis was performed. According to the simple view of reading model (SVR; Gough & Tunmer, 1986), reading fluency was introduced into the model first, and the Herut measure was introduced into the model in the second step. The final model explained 44.2% of the variance in RC ($F(2,1133 = 448.67$, $p < .001$), and both variables contributed significantly to the model. Reading fluency explained 15.3% of the variance in RC,

https://doi.org/10.1017/S0142716422000431 Published online by Cambridge University Press

Table 4. The contribution of the Herut and Elul measures to reading comprehension (combined reading comprehension) ($N = 1{,}133$)

			Model 1			Model 2		
Dependent Measure	Steps	Variables	*B*	*SE B*	*β*	*B*	*SE B*	*B*
RC-combined reading comprehension	1	Reading fluency	0.538	0.038	0.391*	0.180	0.034	.131*
	2	Herut score				0.707	0.029	.597*
		R^2		0.153			0.442	
		R^2 change		0.153*			0.289*	
RC-combined reading comprehension	1	Reading fluency	0.536	0.038	0.389*	0.310	0.036	.225*
	2	Elul score				0.728	0.043	.443*
		R^2		0.151			0.320	
		R^2 change		0.151*			0.169*	

*$p < .001$.

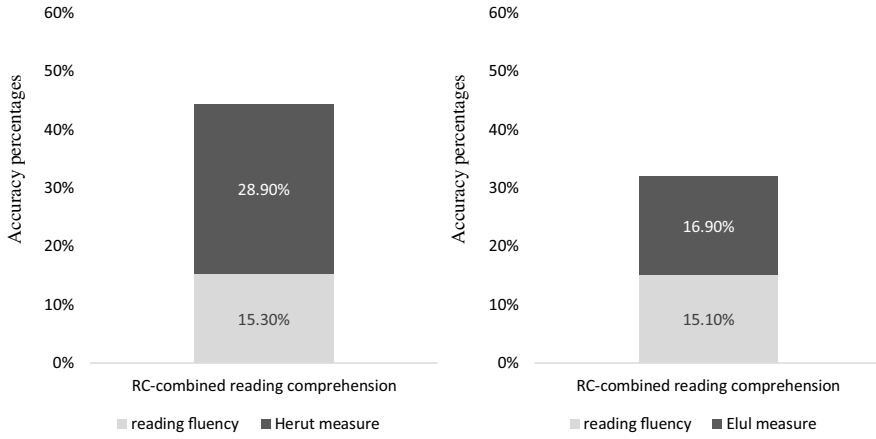

Figure 1. The contribution of the Herut and Elul measure to variability in reading comprehension (combined reading comprehension).

and the Herut measure explained an additional 28.9% of the variance (see Table 4 and Figure 1).

Next, the same regression was conducted with the Elul vocabulary measure. The parallel model examining the contribution of the Elul test to predicting RC showed that both variables, reading fluency and Elul measure, contributed significantly to the model and jointly explained 32% of variance in RC ($F(2,1076) = 266.894$, $p < .001$). Reading fluency contributed 15.1% to RC, and the Elul measure contributed 16.9% to RC (Table 4 and Figure 1). Thus, across combined RC measure

https://doi.org/10.1017/S0142716422000431 Published online by Cambridge University Press

Table 5. The contribution of different types of words (Emotion vs. nonemotion) to reading comprehension in combined reading comprehension ($N = 1{,}133$)

Test type	Regression model	Steps	Variables	*B*	*SE B*	*β*	ΔR^2
RC-combined reading comprehension	Model 1	1	Reading fluency	.161	0.034	0.117*	0.152*
		2	Nonemotion words	.281	0.026	0.304*	0.212*
		3	Emotion words	.369	0.028	0.376*	0.084*
	Model 2	1	Reading fluency	.161	0.034	0.117*	0.152*
		2	Emotion words	.369	0.028	0.376*	0.239*
		3	Nonemotion words	.281	0.026	0.304*	0.057*

*$p < .001$.

(Tamar and Literal and Inferential Meaning), the Herut vocabulary measure explained almost double the percent of variance in RC than did the Elul vocabulary measure.

The contribution of different types of words (Emotion vs. nonemotion) to RC

In order to examine the contribution of the two types of vocabulary words in the Herut test (emotion and nonemotion words) to predicting RC, two hierarchical regression analyses were performed. According to the simple view of reading model, reading fluency was entered in the first step of the regression. Because there is no previous research regarding the unique contribution of emotional vocabulary to RC, two regression models were performed; once when the nonemotion words were entered in the second step, followed by the emotion words, and once when the emotion words were entered in the second step, followed by the nonemotion words.

The findings showed that in both models, all variables, reading fluency, nonemotion words, and emotion words significantly explained 44.8% of the variance in RC ($F(3,1130) = 305.983$, $p < .001$). When the nonemotion words were entered in the second step of the model, they significantly explained 21.2% of the variance in RC. The emotion words, inserted in the third step, significantly added another 8.4% to the explained variance, as can be seen in Table 5 and in Figure 2. In the alternative model, when the emotion vocabulary was entered in the second step, it explained 23.9%, and nonemotion vocabulary was entered on the third step and significantly added 5.7% to the variance in RC, as can be seen in Figure 2 and Table 5.

Analyses clearly demonstrate that nonemotion words and emotion vocabulary are indeed separable constructs and support the categorization of word types put forth in this research. This is because across all regression models, both types of vocabulary knowledge made significant independent contributions to RC.

https://doi.org/10.1017/S0142716422000431 Published online by Cambridge University Press

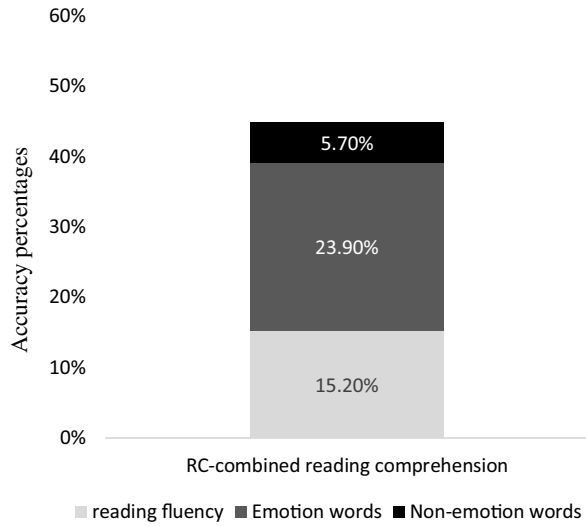

Figure 2. The contribution of different types of words to reading comprehension (combined reading comprehension).

Further, emotion vocabulary makes a slightly greater contribution to explaining variance in RC than does nonemotion vocabulary in integrated RC measure (the Herut and Literal and Inferential measure).

Discussion

Validation of Herut measure

The current study aimed to validate the Herut measure, which evaluates the vocabulary knowledge of fourth- and fifth-grade Hebrew-speaking students. For this purpose, we examined convergent and construct validity. The findings confirm the validity of the Herut measure and thus support the notion regarding different types of vocabulary knowledge emotion and nonemotion words each add unique variance to the prediction of RC. A vocabulary measure that includes both types of words in this study was found to be more valid and useful for predicting RC. In addition, we suggest that assessment on a combined vocabulary measure that includes emotion words will provide a more nuanced approach for developing tailored intervention based on word type (Katzir, 2022, in press).

The current findings also illustrate the strong connection between language and emotion in RC, corroborating and extending Beck's previous findings (Beck et al., 2012). In their study, strong positive correlations were found between linguistic measures, including receptive vocabulary, and emotional abilities among 7- to 9-year-old children. A plausible explanation of these results is that language promotes emotional ability by a process of lexical conceptualization. These findings also reinforce the role of emotional components in RC (Katzir & Primor, in press).

https://doi.org/10.1017/S0142716422000431 Published online by Cambridge University Press

The contribution of different vocabulary measures to RC

After assessing the correlations between the measures and word types, we assessed the contribution of the different vocabulary measures to RC. The findings showed that reading fluency, as well as vocabulary measures (Herut and Elul), together significantly contributed to RC, across two different measures. The overall percent of variability explained in this study is relatively low compared to what is typically observed in predicting RC in other languages (Primor et al., 2011). However, it is similar to previous studies in the Hebrew language. For example, a study that evaluated the contribution of different cognitive and linguistic variables (vocabulary, morphological processing, word reading fluency) to RC among 190 Hebrew-speaking readers in fourth grade, only 25–34% of variance was explained (Primor et al., 2011)

Importantly, because we included reading fluency in the regression models, the analysis controls for its covariability with vocabulary. Nonetheless, the unique contribution of the Herut measure, which includes emotion words and curriculum-based vocabulary, was found to be higher than the unique contribution of the Elul measure across both RC measures.

The greater contribution of the Herut measure compared to the Elul measure may be due to a number of reasons. First, the items included in the Herut measure were selected from students' study materials according to the age group. It seems reasonable to assume that children's mastery of vocabulary from their curriculum may contribute to RC success more than mastery of rare, lower frequency vocabulary. Beyond the fact that the words in the Herut measure were curriculum-based, they also reflected a wider range of words, including emotion words.

Secondly, most of the words (82%) in the Herut measure are categorized as mental words (emotion or cognitive), which are abstract words describing internal processes and situations (Hall & Nagy, 1986; Ravid & Egoz-Liebstein, 2012) and which are acquired at later stages of development (Ravid & Egoz-Liebstein, 2012). Thus, the mastery of mental and abstract words, unlike concrete words, might be more strongly related to the ability to achieve RC in upper elementary school.

Another explanation for the greater contribution of the Herut measure may be related to the type of assignment: The Herut measure assesses vocabulary depth, while the Elul measure is an assessment focused on vocabulary breadth. Thus, a study by Ouellette (2006) demonstrated that for fourth-grade students, vocabulary depth (a definition and synonym task) explained unique variance in RC than vocabulary breadth (the PPVT). Thus, depth of vocabulary knowledge seems to play a stronger role in RC than does vocabulary breadth.

Finally, the greater contribution of the Herut measure may be related to the nature of the task – success in the Herut required being able to understand definitions, which in itself requires RC, whereas the Elul does not require this.

The contribution of vocabulary types to RC

The final aim of the study was to examine the contribution of different types of words from the Herut measure (emotion vs. nonemotion words) to predicting

https://doi.org/10.1017/S0142716422000431 Published online by Cambridge University Press

RC. The results showed a unique contribution of each type to explaining variance in RC. Further, the contribution of emotion words was slightly larger than that of nonemotion words. We offer several explanations for this finding.

First, accuracy rates were significantly higher for emotion words compared to nonemotion words. Students tend to use existing knowledge resources. They have more knowledge about emotion vocabulary and therefore this knowledge is a resource that they can use to support greater RC.

The second explanation is related to the nature of words and their importance. Emotion words convey information about inner states, attitudes, beliefs, social contexts, motives, values, and behaviors (Shiota & Kelther, 2005), thus mediating the inner world of the individual and their environment (Hobbs & Gordon, 2011). During reading, the reader acquires information from the text, and at the same time collects evidence of mental information, such as the emotional states of characters and events. Sentences that describe characters' mental states are fundamental for the interpretation and comprehension of the characters' aims and actions (Graesser et al., 1994) and are central for constructing meaning (Miall, 1988). Identifying and processing emotional states in the text contribute to creating coherence between items, understanding causality, as well as understanding the chain of events and creating a mental model (Graesser et al., 1994; León et al., 2015). Therefore, emotion words may serve as an anchor for making connections between details and events in the text, thus contributing to a better understanding of what is being read, especially in narrative texts.

Another explanation is related to word processing and availability. Emotion words, unlike neutral words, have a value (positive/negative) dimension and are associated with arousal and physiological responses (Hsu et al., 2015; Pavlenko, 2008). Differences between emotion words and neutral words are also evident in information processes. For example, emotion words were found to be identified more quickly and recalled more accurately than neutral words (Kissler et al., 2009; Kissler et al., 2007). Furthermore, eye movement during reading emotion words differs from those observed during reading neutral words (Hinojosa et al., 2010; Scott et al., 2012). It is quite possible that the physiological arousal, memory processes, identification, and processing of emotion words also contribute to RC processes.

Limitations and future directions

There are several limitations to this study. First, the Herut test includes more items classified as emotion words compared to nonemotional words. It is possible that this gap in favor of emotion words may have affected the results.

Second, the data were collected in a single cultural context (Israel), and thus it is currently unknown whether the distinction between emotion and nonemotion words characterizes other societies as well. Thus, it will be interesting to examine whether and to what extent emotional expression in different cultures contributes to RC.

In addition, the vocabulary measures used and compared in this study (Herut compared to Elul) differ along several dimensions. It is difficult to pinpoint the exact reason for the difference manifested in the higher contribution of the Herut

https://doi.org/10.1017/S0142716422000431 Published online by Cambridge University Press

measure. This may be due to the higher frequencies of the words in the Herut measure, and it may also be due to the differences in the task characteristics (definition vs. recognition).

Future avenues of research could include further research on possible differences across types of vocabulary knowledge, such as the distinction between positive and negative emotions. Finally, based on our findings, the contribution of intervention programs that include a rich variety of emotion words should be considered as a promising way of supporting children's development and literacy acquisition.

Conclusions

The current study describes a meticulous procedure for formulating an innovative vocabulary evaluation measure, which includes emotion vocabulary alongside nonemotion words. The study confirms the validity of the measure based on a variety of rigorous analyses, demonstrates its effectiveness in predicting RC, and demonstrates the contribution of emotional vocabulary to RC.

Importantly, the results demonstrate that a measure that includes emotion words and is curriculum-based is more effective in predicting RC, and that emotion words make a unique significant contribution to RC. The study highlights the importance of incorporating emotion vocabulary in assessments as well as in instructional processes and emphasizes the contribution of different types of vocabulary to RC. Emotion words must be recognized as integral components of vocabulary assessment and intervention.

Replication package. Replication data and materials for this article can be found at https://osf.io/2XTRP/

Financial support. This work was supported by a grant from the Israel Department of Education awarded to Tami Katzir, Orly Lipka, Anat Prior, and Michal Shany. This work was also partially supported by a PhD fellowship to TSS from the Israel Scholarship Education Foundation (ISEF).

Conflict of interest. The authors declare that they have no competing interests related to this manuscript.

References

Altarriba, J., & Basnight-Brown, D. M. (2010). The representation of emotion vs. emotion-laden words in English and Spanish in the Affective Simon Task. *International Journal of Bilingualism*, **15**(3), 310–328.

Altarriba, J., Bauer, L. M., & Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods, Instruments, & Computers*, **31**(4), 578–602.

Argaman, O. (2006). *The denotations of emotion words and the lexical means for their expression*. PhD diss. University of Haifa. [in Hebrew].

Beck, I. L., McKeown, M. G., & Kucan, L. (2013). Bringing words to life: Robust vocabulary instruction. Guilford Press, vol. 2 (pp. 19–39).

Beck, L., Kumschick, I. R., Eid, M., & Klann-Delius, G. (2012). Relationship between language competence and emotional competence in middle childhood. *Emotion*, **12**(3), 503–516.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, **34**(2), 213–238.

Coxhead, A. (2011). The Academic Word List ten years on: Research and teaching implications. *TESOL Quarterly*, **45**, 355–362.

Denham, S. A. (2007). Dealing with feelings: How children negotiate the worlds of emotions and social relationships. *Cognition, Brain, Behavior*, **11**(1), 1–48.

Dunn, L. M. (1965). *Expanded manual for the Peabody picture vocabulary test*. American Guidance Service.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody picture vocabulary test-revised*. American guidance service, Incorporated.

Egoz-Liebstein, T. (2009). *Mental verbs in the discourse of children 2–8 years old*. [Unpublished Master's Thesis]. Tel-Aviv University. [in Hebrew].

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska symposium on motivation 1971*, **vol. 19** (pp. 207–282). University of Nebraska Press.

Ekman, P. (1982). *Emotion in the human face*. Cambridge University Press.

Fontaine, J. R., Scherer, K. R., & Soriano, C. (Eds.) (2013). *Components of emotional meaning: A sourcebook*. Oxford University Press.

Frost, R., & Plaut, D. (2005). *The word-frequency database for printed Hebrew*. The Hebrew University of Jerusalem.

Gardner, D., & Davies, M. (2013). A new academic vocabulary list. *Applied Linguistics*, **35**(3), 305–327.

Gendron, M., Lindquist, K. A., Barsalou, L., & Barrett, L. F. (2012). Emotion words shape emotion percepts. *Emotion*, **12**(2), 314.

Gough, P. B. & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, **7**(1), 6–10.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, **101**(3), 371–395.

Graves, M. F., Baumann, J. F., Blachowicz, C. L. Z., Manyak, P., Bates, A., Cieply, C., Davis, J. R., & Von Gunten, H. (2013). Words, words everywhere, but which ones do we teach? *Reading Teacher*, **67**(5), 333–346.

Hall, W. S. & Nagy, W. E. (1986). Theoretical issues in the investigation of the words of internal report. In I. Gopnik & M. Gopnik (Eds.), *From models to modules: Studies in cognitive science from the McGill workshops* (pp. 26–65). Ablex.

Harris, P. L. (1993). Understanding emotion. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotion* (pp. 237–246). Guilford Press.

Hinojosa, J. A., Méndez-Bértolo, C., & Pozo, M. A. (2010). Looking at emotional words is not the same as reading emotional words: Behavioral and neural correlates. *Psychophysiology*, **47**(4), 748–757.

Hobbs, J. R. & Gordon, A. S. (2011). The Deep Lexical Semantics of Emotions. In A. Khurshid (Ed.), *Affective computing and sentiment analysis: Emotion, metaphor and terminology* (pp. 27–34), Springer.

Hsu, C. T., Jacobs, A. M., Citron, F. M., & Conrad, M. (2015). The emotion potential of words and passages in reading Harry Potter – An fMRI study. *Brain and Language*, **142**, 96–114.

Izard, C. E., Fine, S., Schultz, D., Mostow A., Ackerman B., & Youngstrom, E. (2001). Emotion knowledge as a predictor of social behavior and academic competence in children at risk. *Psychological Science*, **12**(1), 18–23.

Izard, C. E., Schultz, D., & Ackerman, B. P. (1996). *Emotional knowledge, social competence and behavior problems in disadvantaged children*. Unpublished manuscript. University of Delaware.

Katzir, T. (in press). The feeling of reading in a changing world. In M. Bishop, S. Sorondo & M. Suárez-Orozco (Eds.), *Education the global impact*, Columbia University Press.

Katzir, T., Lipka, O., Prior, A., & Shany, M. (2019). From mapping to intervention, a multi componential model for Teacher Training. *Report submitted to the Chief of Science, Israeli Ministry of Education* (pp. 1–200).

Katzir, T., Schiff, R., & Kim, Y.S (2012). The effects of orthographic consistency on reading development: A within and between cross-linguistic study of fluency and accuracy among fourth grade English- and Hebrew- speaking children. *Learning and Individual Differences*, **6**, 673–679.

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III*. John Wiley & Sons.

Kissler, J., Herbert, C., Peyk, P., & Junghofer, M. (2007). Buzzwords: Early cortical responses to emotional words during reading. *Psychological Science*, **18**(6), 475–480.

Kissler, J., Herbert, C., Winkler, I., & Junghofer, M. (2009). Emotion and attention in visual word processing – An ERP study. *Biological Psychology*, **80**(1), 75–83.

León, J. A., Dávalos, M. T., Escudero, I., Olmos, R., Morera, Y., & Froufe, M. (2015). Effects of valence and causal direction in the emotion inferences processing during reading: Evidence from a lexical decision task. *Annals of Psychology*, **31**(2), 677–686.

Lewis, M. (1993a). The emergence of human emotions. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotion* (pp. 223–235). Guilford Press.

Lewis, M. (1993b). Self-conscious emotions: Embarrassment, pride, shame, and guilt. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotion* (pp. 536–573). Guilford Press.

Miall, D. S. (1988). Affect and narrative: A model of response to stories. *Poetics*, **17**(3), 259–272.

Miller, A. L., Gouley, K. K., Seifer, R., Zakriski, A., Eguia, M., & Vergnani, M. (2005). Emotion knowledge skills in low income elementary school children: Associations with social status and peer experiences. *Social Development*, **14**(4), 637–651.

NAEP: National Assessment of Educational Progress (2009). *Reading assessment and item specifications for the 2009 National Assessment of Educational Progress*. American Institutes for Research.

NAGB: National Assessment Governing Board (2005). *Reading framework for the 2009 National Assessment of Educational Progress*. American Institutes for Research.

Neuman, S. B., & Wright, T. S. (2014). The magic of words: Teaching vocabulary in the early childhood classroom. *American Educator*, **38**(2), 4–13.

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, **98**(3), 554–566.

Oxford University (2021). Oxford wordlist used in Australian primary schools. Oxfordowl. https://translate.google.com/translate?hl=iw&sl=en&tl=iw&u=https%3A%2F%2Fwww.oxfordowl.com.au%2Foxford-wordlist&anno=2&prev=search

Pavlenko, A. (2008). Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and Cognition*, **11**(2), 147–164.

Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly*, **42**(2), 282–296.

Pearson, P. D., Palincsar, A. S., Biancarosa, G., & Berman, A. I. (Eds.) (2020). *Reaping the rewards of the Reading for Understanding initiative*. National Academy of Education.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, **11**(4), 357–383.

PPVT-4 Publication Summary Form (2007). Retrieved from: http://images.pearsonclinical.com/images/Products/PPVT-IV/ppvt4.pdf

Primor, L., Pierce, M., & Katzir, T. (2011). Predicting reading comprehension of narrative and expository texts among Hebrew-speaking readers with and without a reading disability. *Annals of Dyslexia*, **61**, 242–268.

Ravid, D. & Egoz-Liebstein, T. (2012). Mental state verbs in Hebrew Lexical. *Chelkat Lashon*, 43–44, 241–224. [in Hebrew].

Ricketts, J., Bishop, D. V., Pimperton, H., & Nation, K. (2011). The role of self-teaching in learning orthographic and semantic aspects of new words. *Scientific Studies of Reading*, **15**(1), 47–70.

Ruffman, T., Slade, L., & Crowe, E. (2002). The relation between children's and mothers' mental state language and theory of mind understanding. *Child Development*, **73**(3), 734–751.

Sabag-Shushan, T., & Katzir, T. (2018). *Tamar test for assessing reading comprehension*. University of Haifa.

Sabag-Shushan, T., Yosefi, G., Katzir, T., & Prior, A. (2018). *Herut test for assessing emotion and non-emotion words*. University of Haifa.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, **44**(4), 695–729.

Schmitt, N., Nation, P., & Kremmel, B. (2020). Moving the field of vocabulary assessment forward: The need for more rigorous test development and validation. *Language Teaching*, **53**(1), 109–120.

Scott, G. G., O'Donnell, P. J., & Sereno, S. C. (2012). Emotion words affect eye fixations during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **38**(3), 783–792.

Sedita, J. (2005). Effective vocabulary instruction. *Insights on Learning Disabilities*, **2**(1), 33–45.

Seigneuric, A., & Ehrlich, M. F. (2005). Contribution of working memory capacity to children's reading comprehension: A longitudinal investigation. *Reading and Writing*, **18**(7–9), 617–656.

Shahar-Yames, D., & Prior, A. (2018). The challenge and the opportunity of lexical inferencing in language minority students. *Reading and Writing*, **31**(5), 1109–1132.

Shany, M. & Blicher, S. (2017). *Literal and homiletic meaning. Reading comprehension test for fourth and fifth grade*. Department of Learning Disabilities and Edmond J. Safra Brain Research Center, University of Haifa.

https://doi.org/10.1017/S0142716422000431 Published online by Cambridge University Press

Shany, M., Prior, A., & Blicher, S. (2017). *Vocabulary assessment for elementary school*. Department of Learning Disabilities and Edmond J. Safra Brain Research Center, University of Haifa.

Shatil, E., Nevo, B., & Breznitz, Z. (2007). *Elul test. A standardized diagnostic test for learning disabilities*. University of Haifa.

Shiota, M. N., & Kelther, D. (2005). What do emotion words represent? *Psychological Inquiry*, **16**(1), 32–37.

Silverman, R. D., Proctor, C. P., Harring, J. R., Doyle, B., Mitchell, M. A., & Meyer, A. G. (2014). Teachers' instruction and students' vocabulary and comprehension: An exploratory study with English monolingual and Spanish–English bilingual students in Grades 3–5. *Reading Research Quarterly*, **49**(1), 31–60.

Torgrsen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *TOWRE: Test of Word Reading Efficiency. Examiner's manual*. PRO-ED.

Cite this article: Sabag-Shushan, T., Katzir, T., and Prior, A. (2023). Assessment of emotion word vocabulary and its contribution to reading comprehension. *Applied Psycholinguistics* **44**, 29–50. https://doi.org/10.1017/S0142716422000431