

# THE DEVELOPMENT OF A DIAGNOSTIC TEST ASSESSING YOUNG LEARNER'S KNOWLEDGE OF PHRASAL VERBS

István Thékes

**Abstract:** Vocabulary knowledge is a good predictor of general language proficiency (Boers, Demecheleer, Coxhead, & Webb, 2014). It has been reported that receptive vocabulary knowledge predicts productive vocabulary knowledge (Laufer & Nation, 1999). With the emergence of the lexical approach (Lewis, 1993) in language teaching, a new avenue was opened for vocabulary and phrasal verbs (PV) research. In this study, I will present the attempt to develop a diagnostic test battery assessing the knowledge of PVs. The procedure of the instrument development will be presented, and it will be pointed out that six equivalent tests have been created on the basis of task solving modality and frequency lists. The basis of selecting items for the PV test was Garnier and Schmitt's PHaVE List. Since the new diagnostic test assessing YLs' PV knowledge will first be used amongst Hungarian 6th graders, it was also necessary to check which PVs they were expected to learn) from 4th grade until 6th grade. Similar to Thékes' (2016) procedure, the items were divided into three categories. Three-word categories were established on the basis of the BNC list and the amount of occurrence of a particular word in the course-books. Three perspectives served as the basis of classifying words into categories: 1) word frequency based on the PHaVE List, 2) occurrence of the words in course-books used by Hungarian YLs, 3) personal judgement on the assumed difficulty of the word. The process of determining word categories is presented in the paper. Further test development and future empirical assessment and classroom implications are discussed.

**Keywords:** vocabulary assessment, phrasal verbs, young learners, diagnostic testing, new instrument development

## 1. Introduction

The 1990s saw an increasing number of studies focusing on FL vocabulary learning and the literature has been growing ever since then by extending the knowledge on such areas as FL vocabulary assessment (Laufer, Elder, Congdon, & Hill, 2004; Nation, 2001) the FL mental lexicon (Singleton, 1999; Zareva, 2007), corpus studies (Kilgarriff, 1997; Nation & Macalister, 2010). It has also been affirmed that vocabulary knowledge is a good predictor of general language proficiency (Boers, Demecheleer, Coxhead, & Webb, 2014). It has been reported that receptive vocabulary knowledge predicts productive vocabulary knowledge (Laufer & Nation, 1999). With the emergence of the lexical approach (Lewis, 1993) in language teaching, a new avenue was opened for vocabulary research. An expert on language teaching uncompromisingly concludes "Without grammar very little can be conveyed, without vocabulary nothing can be conveyed" (Thornbury, 2002, p. 18). Vocabulary is also considered as one of the strongest predictors of FL proficiency (Schmitt, 2008). Significant correlations have been found between receptive vocabulary knowledge and FL reading comprehension (Qian, 2002; Schmitt, Jiang & Grabe, 2011).

Having seen the arguments on behalf of vocabulary teaching, it must also be pointed out that the learning and teaching of phrasal verbs (PVs) is gaining popularity in the FL learning literature. Since PVs make up a considerable part of the vocabulary used by native speakers in everyday context, more and more attention are turned towards the involvement and teaching of PVs in the foreign language

(FL) curriculum. Several studies (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Gardner & Davies, 2007; Garnier & Schmitt, 2015) have been published in the context of PV research.

In this study, I will present the attempt to develop a diagnostic test battery assessing the knowledge of PVs. The procedure of the instrument development will be presented, and it will be pointed out that six equivalent tests have been created on the basis of task solving modality and frequency lists.

## 2. Young learners

Even though the definition of ‘young learners’ (YLS) might have variations across the globe, this term is entirely clarified in this article. An agreement exists among the European Union member states that children before their age of six are called ‘very YLS’. From age seven they are termed ‘YLS’. Nikolov and Mihaljevič Djigunović (2006) clearly state that learners can be considered young up to the age of fourteen. In certain contexts, the definition of ‘YLS’ allows an age extension only to age twelve.

In the context of the current research, the sample comprises 6th graders at the age of twelve. The majority of 6th graders in Hungary start learning English as a foreign language (EFL) as a mandatory subject at the age of ten when they are 4th graders. Some parents motivate their YLS to begin their English language studies as early as kindergarten or the first years of primary school. However, the assertion can be made that most of the learners involved in the research have an experience of two years of English learning. They are termed young learners throughout the study and the abbreviation ‘YLS’ will be used to indicate this age range.

## 3. Context of developing the diagnostic test assessing PV knowledge

Hungarian 6th graders are 12-year-old learners, most of whom, have three or four 45-minute lessons in English in primary schools. The majority of the public schools do not provide more than four English lessons for students per week (Fazekas, 2009). However, most of the students in Hungary have more exposure to English than the three or four occasions determined in the school curriculum. They attend private language lessons or courses organized by local language schools in the afternoons and they are exposed to a large amount of English by using the Internet.

Besides being exposed to vocabulary learning in school and private lessons, students are also believed to learn vocabulary by listening to songs on Youtube and reading posts on social media sites. In the schools where I conducted the pilot study I interviewed the teachers to make sure I am aware of what course-books were used. Information was provided that they used course-books published by either of the three major publishers: Cambridge University Press, Oxford University Press and Pearson. It is characteristic of these course-books and workbooks that they have well-designed studies and contain a great deal of visual material. As for course-books, the investigation of how teachers apply the course-books in practice and what methods they use to teach English is also an important aspect. In an empirical study conducted by Nikolov (2008), it was pointed out that the observed teachers in the research tended to use the grammar-translation method (Harmer, 2012), and their native Hungarian to explain grammar and vocabulary meaning. It was asserted that English words were generally taught with their Hungarian equivalents and very few motivating techniques are applied to help learners learn vocabulary.

Until the 1990s, students in primary school started learning Russian in 4th grade and normally they attended three Russian lessons a week until 12th grade, the final grade in public education. Because of the difficult orthography and lack of willingness on the part of the Hungarian learners to identify themselves with the language, Russian was not a popular subject and most of the learners left public education without being able to communicate in this language (Nikolov, 2007). Since Hungary became a democratic country (the first democratic elections were held in 1990) Russian teachers have been retrained into English teachers. Learning English and German has become popular, especially English as lingua franca, since the turn of the millennium (Csizér & Dörnyei, 2005). Nowadays, an increasing number of children start their FL studies prior to the mandatory age of ten. Due to parental pressure, more and more YLS begin to study English before grade 4 (Nikolov & Szabó, 2011). The most popular FL is English but German, French, Italian, Spanish and Chinese are also offered in some schools.

Hungarian YLs gain access to English words from three main sources: (1) public school classroom, (2) private lessons, and (3) incidental instances of hearing or reading words. As for the classroom, vocabulary input can be received by the student from teacher talk. One study puts English education in Hungary into focus (Nikolov, 1999). A lot of useful observations can be made based on this study. One is that teachers in Hungary use NL in the majority of the classroom time and tend not to use pictures, videos or songs to teach language and vocabulary, a method that learners would prefer according to their report. On the basis of Nikolov's (1999) data originating from a series of classroom observations in Hungary, I reckon that English words are predominantly taught with the use of NL. I observed lessons prior to the pilot study of the vocabulary test and saw the techniques of grammar-translation method used in the classrooms. Words were basically taught with one technique: an unknown word arose from context and the teacher gave the Hungarian meaning. There is a likelihood that the lack of variety in teaching techniques limits the chances of the learners to learn words rapidly and efficiently in the classroom.

#### 4. The use of corpora in vocabulary research

Corpus linguistics is a rapidly developing field of applied linguistics. A large amount of corpora are being developed all over the world for a lot of languages and for a lot of jargons also. While the compilation of different corpora had been a gigantic and imprecise effort before computers, nowadays exact data can be gathered with relative efficiency. This efficiency is really relative since the common endeavor of Cambridge University and the University of Nottingham, the CANCODE spoken corpus of British English took eight years to finalize by transcribing and coding five million words. Besides numerous English corpora, an attempt has been made to assemble corpora in most languages.

Frequency is the most underlying concept that is examined in corpus linguistics. The most elementary thing that can be deduced from studying the language in a corpus is how many times a particular word occurs. The earliest corpora in research gave the frequency of a word as the first piece of information to researchers.

The General Service List (West, 1953) and University Word List (Xue & Nation, 1984) were outlined with the aim of measuring lexical richness in a new manner. This profile was then called the Lexical Frequency Profile (LFP) and was developed by Laufer and Nation (1995). The authors claimed they had the intention to exclude subjective judgments in the assessment of writing quality and it was efficient in the measurement of how vocabulary size was reflected in use. Laufer (2001) also claims that the profile has no topic dependency and is a reliable measure as long as the topic is general, and the writers are not required to apply any jargon. Jargon, in its nature, implies that a large amount of low-frequency words is used. The profile validated by these two researchers is not suited for assessment of any specific jargon.

##### 4.1 Widely used corpora

In this section two corpora are presented in details from various perspectives such as an insight into their original purpose, their contents, and their area of usage. It must be noted that five different corpora are generally used for vocabulary research: (1) British National Corpus (BNC), (2) Contemporary Corpus of American English (COCA), (3) American National Corpus (ANC), (4) Cambridge and Nottingham Corpus of Discourse in English (CANCODE), and (5) the Child Language Data Exchange System (CHILDES).

Since the BNC and the COCA were considered for use in the selection of words during test development, I elaborate on these two corpora. The BNC is available at <http://www.natcorp.ox.ac.uk/>. The development process of the BNC was published by Kilgarriff (1997). This corpus was launched in the early 1990s and was applied as a basis of vocabulary assessment to a large extent. Work began in 1991 and the first version was available for public use in 1994. It is the most cited corpus in word knowledge assessment and it is generally the basis of word selection in the development of diagnostic vocabulary texts. The BNC is considered as a main source for anybody involved in language teaching. It contains more than 100 million words and has a large part of spoken corpus.

The COCA is available at <http://corpus.byu.edu/coca/>. It is claimed to be the largest freely-available corpus of English on its website. This corpus was developed and is under constant construction by Mark Davies of Brigham Young University. It is described as the first large, genre-balanced corpus of any language (Davies, 2010). The latest update was made in 2012. It contains 400 million words and is used broadly by researchers, linguists, teachers and translators. On the COCA website, the user should simply type in the word and the site generates all the necessary information (frequency, frequency rank) in less than a second.

## 5. Phrasal verb frequency lists

Phrasal verbs (PVs) are almost unanimously defined in the literature (Merriam-Webster, 2016; Cambridge Dictionary, 2016) as a phrase (as ‘*take off*’ or ‘*look down on*’) that combines a verb with a preposition or adverb and that functions as a verb whose meaning differs from the combined meanings of the individual words.

Similar to the BNC and the COCA, attempts have been made to create a reliable pool of phrasal verbs based on frequency. As Garnier and Schmitt (2015) note frequency of occurrence is an appropriate indicator of usefulness. There are variations in the estimation of PVs. McCarthy and Dell (2004) state that there are over 5,000 PVs. Gardner and Davies (2007) go as far as stipulating that there are a total of 12,508 PVs in English. This is an indication of the need to set up frequency lists of PVs.

Gardner and Davies (2007) used the BNC as the only data-source to establish the list of the most frequent PVs. They composed the inventory of the most frequent PVs. They asserted that the top 20 lexical verbs that one finds in PV constructions make up 53.7% of all PVs; moreover, they also pointed out that these 20 lexical verbs account for half of the PVs in the BNC when one combines them with only eight particles. It must be noted that Gardner and Davies (2007) emphasized the highly polysemous nature of PVs by pointing out that lexical items on their list have a 5.6 meaning sense average figure.

Liu (2011) endeavored into establishing a brand-new inventory of the most frequent PVs by reviewing previous lists including Gardner and Davies (2007). Liu (2011) examined 8,847 PVs on the basis of frequency and ended up incorporating 150 PVs into the list. According to him, there is no significant difference in the use of PVs between British and American English and in contrast to the continuously growing number of words, PV use has remained relatively unchanged over the past decades.

The most recent list of PVs called PHaVE List was compiled by Garnier and Schmitt (2015). First and foremost, they took Gardner and Davies’ (2007) and Liu’s (2011) list into consideration when analyzing the included PVs. They considered all the PVs used by Gardner and Davies (2007), plus 48 other items extracted by Liu from the COCA. Thus, their list added up to 150 words. Garnier and Schmitt (2015) sincerely admitted to knowing 150 PVs is not sufficient considering the enormous amount of PVs in English. However, they made the decision that they intended to make the list as useful as possible for learners and teachers of English. So, their limitation of the list to no more than 150 words was supported by a pedagogical perspective motivated by practicality. They also wanted to avoid the list being too long, which is a prerequisite for a frequency list to be truly meaningful (Liu, 2011). They further suggested that learning only these PVs would be beneficial for the students (Garnier & Schmitt, 2015). As for what information to provide in their PHaVE List, they decided to give the following information: 1) meaning and the connotations of these words; 2) meaning sense frequency percentages; and 3) example sentences. The ordering of the items, similar to Liu’s (2011) was made by frequency order. The sources they used to compile the PHaVE list besides Gardner and Davies (2007) and Liu (2011) were well-known and established dictionaries (e.g., Oxford Dictionaries, Merriam-Webster, Collins COBUILD Phrasal Verbs Dictionary, etc.) and a lexical database (WordNet Search 3.1). They noted that PV dictionaries include more sophisticated and refined distinctions than general dictionaries. As for the corpus they chose, it was the COCA since it offered the researchers the four following advantages: 1) it is very large; 2) it is balanced across several genres; 3) it is frequently updated; and 4) it is freely accessible. Following the corpus analysis procedure and an inter-rater reliability procedure, the final list was compiled. An example is provided here with the lexical item ‘work out’ (Table 1).

**Table 1. Example of the word 'work out' in the Granier and Schmitt (2015) PHaVe List**

Meaning sense	Sample sentence
Plan, devise or think about STH carefully or in detail (33%)	We still need to work out the details of the procedure.
Exercise in order to improve health or strength (23%)	He works out at the gym 5 times a week.
(+well/badly) Happen or develop in a particular way (15%)	Everything worked out well in the end.
Prove to be successful (12.5%)	Despite our efforts, it just didn't work out

## 6. Assessing vocabulary

### 6.1 Foreign language vocabulary tests

Ever since vocabulary came into the focus of foreign language learning studies, assessment of word knowledge has been perceived as a fundamental issue in the research of this domain. This study provides an insight into how vocabulary is assessed and what types of validated and reliable instruments exist in the literature. Apart from presenting these instruments, special attention will be drawn to (1) the computerized versions of these instruments as in the 21st century diagnostic assessment is predominantly executed in an online environment (Laufer et al., 2004) and (2) whether the data collection instruments to be discussed have versions designed for YLs. It must also be highlighted that there is a consensus among scholars in foreign language vocabulary assessment that various modalities (see more in Laufer et al., 2004) of item assessment exist. Laufer et al. (2004) claim that words may be measured from two perspectives: (1) the form-focused perspective that implies that the test-taker is able to retrieve the form of the word evidencing productive knowledge, (2) the meaning-focused perspective that entails the test-taker can retrieve the meaning of the word evidencing receptive knowledge. Laufer et al. (2004) refer to the productive-receptive dichotomy as active-passive knowledge. Four degrees of knowledge of meaning are discerned, on the basis of two dichotomous distinctions: providing the form for a given concept vs. providing the meaning for a given form; and recall vs. recognition (of form or meaning). These distinctions entail the following four modalities constituting a hierarchy of difficulty: (1) passive recognition that encapsulates recognizing an item in e.g. a multiple choice test, (2) active recognition that encompasses a given definition and four items; in this modality the definition must be matched with the pertaining item, (3) passive recall that incorporates a sentence and the synonym of one item in the sentence must be given by the test-takers, and (4) active recall that comprises a description of items and the initial letters of the items are provided; test-takers are expected to produce the word. In a review article Schmitt (2014) uses different terms for the same concepts. Passive recognition is termed meaning recognition; active recognition is named form recognition whereas passive recall is termed meaning recall and active recall is called form recall. In an attempt to provide instances, sample tasks will be given subsequently. In this article Schmitt's (2014) terminology will be utilized since the passive-active dichotomy is rather obsolete in the current literature (Webb, 2008). Table 2 presents sample tasks of each of the four modalities.

Table 2. Sample tasks of the four modalities

	Recognition	Recall
Meaning	<p>large a) small b) tiny c) huge d) weak</p> <p>Instruction: circle the equivalent of 'large'.</p>	<p>When something is large, it is _____.</p> <p>Instruction: finish the sentence with adequate words</p>
Form	<p>something that is very big in size. a) large b) angry c) hungry d) strong</p> <p>Instruction: circle the word that suits the definition.</p>	<p>Something is very big in size . l_____</p> <p>Instruction: finish the sentence with a word that starts with the given letter.</p>

## 6.2 The Receptive Vocabulary Levels Test

The Receptive Vocabulary Levels Test is simply referred in the literature to as Vocabulary Levels Test (VLT). It operates with a discrete point measure. It requires meaning recognition. The test was developed by Nation (1990) and it was validated by Schmitt, Schmitt and Clapham (2001). Words are selected from such corpora as British National Corpus (Kilgarriff, 1997) and the CANCODE (Cambridge and Nottingham Corpus of Discourse in English) up to five levels: the first 2,000, 3,000, 5,000 and 10,000 most frequent words. These levels bear importance from a research-based perspective. The 2,000-3,000 levels contain high-frequency words whose knowledge is necessary for everyday communication. The 5,000 level is the minimal size which learners can conceive authentic texts. The 10,000 level, contains the most common low-frequency words (Webb, 2010). The fifth level is not grounded on any corpus but includes items from the University Word List (Xue & Nation, 1984).

The test-taker sees six words on the left-hand side and three definitions or synonyms on the right-hand side. They are expected to match the right-hand side items with three of the six words on the left-hand side. This means that the task contains three distractors. In the entire test each level comprises six clusters of six words. Table 3 presents one sample task of the VLT.

Table 3. Sample task of the VLT (Schmitt, Schmitt &amp; Clapham, 2001)

**Instruction: match three of the words from 1) to 6) with three definitions A) - C)**

1 bitter	
2 independent	very small
3 lovely	beautiful
4 merry	liked by many people
5 popular	
6 slight	

Since the test gives estimates of vocabulary size at 5 levels, it can be applied for placement purposes and for diagnosis of vocabulary gaps. Four parallel test versions were developed. The criterion of the development of the test was that the definitions are succinct; the test could be completed in the fastest possible time and with the appropriate arrangement of the possibility of blind guesses could be diminished. In the online version of the VLT the test-taker is expected to write the listed six words next to the three definitions. The evaluation of the test is automatically completed. With the modified version of the online test, Vocabulary Online Recognition Speed Test (VORST) the speed of word recognition can also be examined (Laufer & Nation, 2001).

A version of the VLT designed for YLs has also been developed. Jimenez Catalan and Terrazas Gallego (2008) used the YL version of the instrument with young Spanish YLs of English. They modified the word selection process by involving such low-frequency words as names of animals (e.g., 'lion', 'ostrich', 'tiger') that YLs might know better than high-frequency words used by adults

(e.g., 'beer', 'office', 'wine'). The researchers reported that the YLs' version of the VLT proved to be a valid measure of vocabulary assessment.

### 6.3 Productive Vocabulary Levels Test

With regard to productive knowledge of vocabulary, Laufer and Nation (1995) developed an instrument that measures productive word knowledge. The test took its name after the VLT and the adjective 'productive' was added so that the type of test would be clearly discerned. The test requires form recall on part of the participants. Similarly to the Vocabulary Levels Test, the tasks are divided into frequency clusters: 2,000, 3,000, 5,000, 10,000. In this test sentences are seen by students. In each sentence only the first two or three initial letters of one word are provided. Students must write the missing part of the word. This test is originally named the Test of Controlled Productive Ability (TCPA), nowadays it is referred to as Productive Vocabulary Levels Test (PVLТ). A part of the instrument is presented in Table 4.

**Table 4. Productive Vocabulary Levels Test (Laufer & Nation, 1999)**

**Instruction: Complete the words by filling in the gaps with the proper letters**

---

He likes walking in the fo..... because the trees are beautiful there.

He takes cr.....and sugar in his coffee

The actor took the st..... to perform in the long-awaited play.

---

It is obvious from Table 4, that the sentences following one another are unrelated. The test format resembles a C-test to some extent. In the pilot study of the instrument it was reported by the researchers that the selection of the target words was determined with the aim of avoiding any ambiguity of the meaning of the words. Similar to Schmitt et al. (2001) four test versions were developed. It is worth noting that the test has been criticized from a construct validity point of view. It was pointed out by Read (2000) that the instrument is unlikely to assess productive word knowledge. He argues that some of the items demand only recognition and some of them need more contextual clues than others, thus he is dubious whether the test assesses what it is meant to assess.

### 6.4 Vocabulary Size Test

The Vocabulary Size Test (VST) was developed and validated by Nation and Beglar (2007). It assesses the knowledge of the 14,000 most frequent English words. It implies the modality of meaning recognition similar to the VLT. One sentence is given in each task and one word is underlined in the sentence. Under the sentence four possible options are provided in a multiple-choice format and the test-takers must settle upon which word is interchangeable with the underlined word. The test is available in online version and it renders the assessment of receptive vocabulary rapid and effective. It is a very similar test format to the one applied on the TOEFL test which is also taken in a computerized environment. Table 5 presents a sample task of the VST.

**Table 5. Sample task of the VST (Nation & Beglar, 2007)**

**Instruction: Choose one proper word from items a) - d) that best fits the word in bold**

---

DRIVE: He drives fast

- a. Swims
  - b. Learns
  - c. throws a ball
  - d. uses a car
- 

## 7. Foreign language vocabulary tests for YLs

### 7.1 Principles of designing FL vocabulary tests for YLs

Although most of the above-mentioned diagnostic data collection instruments have been originally designed to assess university students or adults, there have been studies reporting on the testing of

YLS' word knowledge as well. Diagnostic assessment of YLS' FL proficiency and word knowledge empowers teachers with a lot of classroom implications (McKay, 2006).

It is typical of YLS that they use memorized chunks (Thákes, 2016). Their knowledge is implicit in this sense; explicit learning ability that enables them to comprehend rules emerge around adolescence (Nikolov & Szabó, 2011). Most of the YLS learn words rapidly (Orosz, 2009); nevertheless, after they are capable of recognizing words, the ability to use connotations, shades of meaning, synonyms and antonyms is only learned as a result of a long process of learning (Cameron, 2004, p. 32). Three fundamental facts have also been emphasized in the literature: 1) until the age of twelve students know only a limited (not more than 600-700) amount of words in an FL (Laufer, 1997), 2) students hardly ever know the connotations (Schmitt, 2008), and 3) YLS have limited awareness of the derivative forms of a word (Schmitt & Zimmerman, 2002).

Before presenting the findings of studies assessing the word knowledge of YLS, I will elaborate on the characteristic traits and principles of diagnostic testing of FL in the context of YLS. Nikolov and Szabó outlined the principles of diagnostic testing of YLS (2011). These principles are based on the study by Alderson (2005). I will make an attempt to synthesize these principles which, I believe, are the most relevant from the perspective of vocabulary assessment of YLS.

- 1) the purpose of diagnostic tests is to identify the strengths and weaknesses of learners,
- 2) diagnostic tests must result in the treatment of difficulties arising during the learning process,
- 3) diagnostic tests must make it possible to analyze the score of each item in detail and to report the results; thus, they provide feedback in detail and further steps can be taken,
- 4) diagnostic tests are low-stakes tests or bear no consequences so optimal achievement is not hindered by anxiety or any other affective factor,
- 5) diagnostic tests must take into consideration research on FL learning and in a wider sense the results of applied linguistics research,
- 6) diagnostic tests are more likely to be discreet-point tests than integrative, i.e., they focus on certain linguistic elements rather than on global abilities,
- 7) diagnostic tests are more likely to be less authentic than any other level-testing instruments;
- 8) diagnostic tests are more likely to focus on 'lower-level' linguistic abilities than on 'higher-level' abilities,
- 9) diagnostic tests assessing linguistic skills (listening, speaking, reading, writing) are more simple to develop than ones assessing grammar skills,
- (10) diagnostic testing is probably made more efficient by using a computerized platform.

Jang (2014), whose claims can also be subjected to FL vocabulary assessment, goes further by making the subsequent claims concerning the diagnostic assessment foreign language: assessment should (1) be cognitively rich enough to elicit knowledge and skills, (2) measure essential core skills, (3) promote positive learning and assessment experiences, (4) provide consistent and reliable information on proficiency, (5) promote students' ability to self-assess, (6) provide the support needed.

The two sets of principles laid out by Nikolov and Szabó (2011) and Jiang (2014) overlap in several points. It can be concluded that diagnostic vocabulary assessment of YLS ought to (1) give constant feedback, (2) provide information constantly to promote the learning process, and (3) remain low-stakes so that it can provide positive learning experiences.

## 7. 2 Diagnostic online English and German receptive vocabulary size test for YLS

Most recently a FL vocabulary test has been developed and validated by the researchers of the University of Szeged (Vidákovich et al., 2013). The instrument is designed and calibrated to measure diagnostically the vocabulary size of 5th and 6th graders learning English and German as a FL. The selection of the target items was done on the basis of frequency lists and corpora and the test is unique in the sense that the words incorporated in the test are similar in the two languages. The instrument has a multiple-choice test format in that the students see one picture and four words on the screen and they have to decide which word is described by the picture. Unlike the Peabody Picture Vocabulary Test



(Dunn & Dunn, 2007) where only one word matches one picture in one task, in this test there is a likelihood that all four words match the picture or only one word can be matched with the picture; thus test-takers do not automatically exclude any correct item after solving one. The test-takers must click on the buttons next to each word and settle upon whether there is a match or not. The pictures are either simple or complex pictures and students must use identification or implication to figure out the correct answer. The test demands meaning recognition. The instrument has three versions in both languages. The instrument has been applied in an online environment on the eDia platform developed by the ICT specialists of the Institute of Educational Science at the University of Szeged (Molnár, 2013). The test-taking period is short as it takes around ten minutes and apart from the test scores, background data can be processed immediately after the completion of the data collection instrument. Table 6 presents one task of the test.

**Table 6. Example of an item containing a simple picture (Vidákovich, Vígh, S. Hrebik, & Thékes, 2013)**

**Instruction: Choose from words a) – d) that best fit the picture on the left.**



- a) chair
- b) plant
- c) table
- d) theatre

### 7.3 Complex online diagnostic test for young learners (CONDITEYOLE)

Thékes (2016) developed a complex online vocabulary test. Detailed information on the CONDITEYOLE is provided in Thékes' doctoral dissertation (2016) and further in Thékes (2015a, 2015b). Table 7 presents the structure of the test.

**Table 7. Tasks in the diagnostic vocabulary test battery**

Task	Receptive/ Productive	Language skill(s) and modality required (Schmitt, 2014)
1 Choose one proper phrasal verb from items that best fits the word in bold	Receptive	Reading / Meaning recognition
2 Read the definitions and match them the phrasal verbs	Receptive	Listening / Form recognition
3 Match 6 written words with 3 pictures	Receptive	Reading / Meaning recognition
4 Match written words with picture	Receptive	Reading / Meaning recognition
5 Match written definitions with words	Receptive	Reading / Form recognition
6 Write word next to picture	Productive	Writing / Form recall

First, a paper-and-pencil validation study was conducted in 2013 and based on the validation, the final test format was decided on. Data on the online test were collected by Thékes (2016) in 2014 with the participation of nearly 300 YLs.

## 7.4 Conceptualizing the new complex online phrasal verbs test for young learners (COPHAVE Test for YLS)

When conceptualizing the newly developed test assessing YLs' phrasal verbs knowledge, I took into consideration what Alderson (2005), Nikolov and Szabó (2011), Schmitt (2014), and Jang (2014) asserted concerning diagnostic language tests for YLs. The following stipulations received most of the focus: 1) the PV test ought to measure core skills, 2) it must promote positive learning and assessment experiences, 3) it must assess PV knowledge in different modalities, and finally 4) it must be cognitively rich enough to elicit knowledge.

Based on the empirical experiences procured in the online testing of YL's vocabulary with a validated complex test that assess vocabulary in different modalities, it was worthwhile assessing YLs with different types of task because variations in the organization of FL word knowledge could be revealed, plus correlations amongst the different tasks could be analyzed as well. Gaining experience and evidence from the analysis of the results of the CONDITEYOLE, I decided to create three tasks in the COPHAVE Test for YLS: 1) a receptive task in meaning recognition, 2) a receptive task in form recognition, and 3) a productive task in form recall modality.

This decision was made since I surmised that two receptive tasks in two different modalities and one productive task in form recall modality would be sufficient to assess diagnostically YLs' PV knowledge. Table 8 presents the structure of the diagnostic PV test for YLs.

**Table 8. Tasks in the COPHAVE Test for YLS**

Task	Receptive/ Productive	Language skill(s) and modality required (Schmitt, 2014)
1 Choose the phrasal verb that matches the meaning of the verb in bold	Receptive	Reading / Meaning recognition
2 Choose the correct preposition after the verb	Receptive	Reading / Form recognition
3 Write the proper preposition into the gaps	Productive	Writing / Form recall

## 7.5 Selecting the items for the COPHAVE Test for YLS

First and foremost, I contacted Garnier and Schmitt's PHaVE List (2015) upon compiling the items for the CONDITEYOLE. This decision was supported by three main reasons: 1) there is no more profound and carefully compiled list of PVs than that of the two researchers; 2) the list was created from a pedagogical perspective; 3) the list comprises the most frequent PVs in English, several of which YLs can be expected to know.

Garnier and Schmitt (2015) go into detail about the polysemous nature of PVs. They argue that no-one knows anything whether students are aware of the different meaning senses of the polysemous PVs. Polysemy is problematic in the assessment of PV knowledge, especially in case of YLs since they are not supposed to know synonyms and multiple connotations of PVs. In this respect, I made the decision to include exclusively the most frequent meaning sense of the PVs.

Selecting the items from the PHaVE List would have been hardly sufficient. Course-books that YLs use had to be consulted and lessons had to be observed to explore the amount of PV knowledge that can be expected from the YLs. Since the new diagnostic test assessing YLs' PV knowledge will first be used amongst Hungarian 6th graders, it was necessary to check which PVs they were expected to learn from 4th grade until 6th grade. Four teachers of English in four different Hungarian primary schools were called upon to list the course-books they use with their students that start learning in 4th grade. Plus, they were also requested to underline the PVs that they have taught, or the students are likely to know during the three years of the YLs' English studies. Apart from their help, I also checked all the course-books listed by the four teachers and made a list to myself which I eventually compared with the lists of the four teachers. In case a PV was listed in at least three of the five lists (four teachers plus me), it was included in a final list. This inventory comprised 88 PVs.

In spring 2016, I also observed eight English lessons of 6th graders in four primary schools in Szeged, Hungary and paid attention to which PVs the teachers mentioned in their speech, may it be part of an instruction and I also listened to the students' speaking and took note of the PVs that they used verbally in class. Based on classroom observations, another list was thus created that included all the PVs mentioned in class by either teachers or students. This inventory comprised 42 PVs.

At this point I had three lists at my disposal: 1) Garnier and Schmitt's PHaVE List (2015), 2) the list compiled by four teachers and me, 3) the list grounded on classroom observations. The three inventories were compared and any overlap amongst the latter two lists and the PHaVE list resulted in including that particular item in the ultimate list that would serve as the pool of items for the test battery. This procedure led to 47 PVs. Following this I checked the PHaVE List again and completed the ultimate inventory with the seven most frequent PVs on the PHaVE List that were not included amongst the 47 PVs. Thus, the inventory eventually comprised 54 items.

Similar to Thékes' (2016) procedure, the items were divided into three categories. Three-word categories were established on the basis of the BNC list and the amount of occurrence of a particular word in the course-books. The necessity of creating categories is underlined by the fact that major vocabulary tests (Nation, 2001, Laufer & Nation, 1995) include items selected on the basis of layered word list. Three perspectives served as the basis of classifying words into categories: 1) word frequency based on the PHaVE List, 2) occurrence of the words in course-books used by Hungarian YLs, 3) personal judgement on the assumed difficulty of the word. The process of determining word categories is presented below.

Every PV in the test was given a difficulty index calculated from the sum of the three perspectives. Points were given on a scale of one to three based on the perspectives of classifying the PVs. In terms of each perspective a minimum of one point and a maximum of three points were given to the PVs. One point was the indication of easiness and three points were that of difficulty. Personal judgement was done prior to consulting the PHaVE List and course-book occurrence so that prejudice would be avoided. In case I assumed a word easy, I gave it one point; I gave two points to a word I assumed of average difficulty and three points were given to the words that were supposed to be the most difficult. I conducted my judgement on the basis of fifteen years of experience of teaching EFL. As regards the frequency perspective, the word was given one point if it was among the 2,000 most frequent PVs in the PHaVE List, it was given two points if it was between the 2,000 and the 4,000 most frequent words. In case it was outside the 4,000 most frequent PVs, it was given three points. As concerns course-book occurrence, I consulted the course-books used by the students and investigated my item pool with a focus on how frequently the words appear in the books. I wrote ticks next to the words on my list. Afterwards I counted the ticks and gave points to the PVs in the following way: one point to more than six ticks, two points for a number of ticks between three and five, and finally three points for ticks between one and three.

From the process described above it is clear that each word could be given the minimum of three points and the maximum of nine points. The summed points were considered the difficulty indices of the PVs. Based on these difficulty indices, the categories of the PVs were determined. Category 1 contains the simplest and Category 3 contains the most difficult items. Table 9 presents the determination of the categories.

**Table 9. The categories of the words based on index points**

Index points	3-5	6-8	7-9
Category of the word	1	2	3

As mentioned above the final inventory of PVs comprise 54 words. The goal was to create the three categories with equal number of items so that in the analysis of the results the knowledge of the words of different categories could be compared. Table 10 presents the PVs that will be used as items in the tasks of the COPHAVE Test for YLS. The category is also indicated. The list is in alphabetical order.

Table 10. The PVs and their category based on index points

Phrasal Verb	Category
break out	2
bring back	2
build up	3
clean up	3
close down	3
come back	1
come down	2
come in	1
come on	1
come out	1
fill out	2
find out	1
get back	1
get down	3
get in	3
get out	1
get up	1
give in	3
give up	1
go ahead	2
go back	1
go off	3
go on	1
go out	1
go up	2
grow up	1
hang up	3
hold up	2
keep on	2
look around	2
look back	2
look out	2
look up	1
move in	3
move on	2
move out	3
pick up	1
put back	3
put on	3
set out	3
show up	1
shut down	3
shut up	2
sit down	1
slow down	3
sort out	3
stand up	2
start out	3
step back	2
take off	1
throw out	3
wake up	2
walk out	2
write down	2

## 8. Creating the COPHAVE Test for YLS

As presented in Table 10, eighteen items were classified into either of the three categories. As discussed above, a decision was made to create three different tasks in the test battery. Six equivalent tests were developed, each comprising three tasks of the same format. Each task in each test consists of nine items. This means that a test comprises 27 PVs. One PV occurs in three tests out of the six but always in a task that requires knowledge of a different modality. Thus, every PV is assessed in all of the three modalities (meaning recognition, form recognition and form recall). This makes it possible to compare the achievements in the three modalities both at item and at student level.

As pointed out, a 27-item test consists of three tasks and each task involves nine items. Students can reach nine points in each task; thus, their maximum possible achievement is 27 points in each test. When composing the tasks, I made sure that every task would contain equal number of items of the three categories. In each task three items were selected from Category 1, three items were selected from Category 2, and three items were selected from Category 3. This way, not only the task-solving modalities can be compared but the categories as well. In sum, equivalency among the tests was assured by selecting equal number of items of each category and by assessing each PV in each modality. Appendix 1 presents one test out of the six.

As for further principles taken into account, three important aspects of the test battery need to be mentioned: 1) the sentences in the tasks were phrased in a context familiar to 6th graders and motivating for them, which is an important tenet of language tests for YLs (Nikolov, 2011), 2) in the two receptive tasks, the items to be selected are listed in alphabetical order so that students will not surmise any hidden trace behind the order of listed items; thus validity is strengthened, and 3) the sentences are short but interpretable so that the proper PV can be elicited. Furthermore, the format of Task 1 resembles that of the VST (for description, see above), that of Task 2 takes the model of the diagnostic online English and German receptive vocabulary size test for YLs except that no picture is applied, and Task 3 is basically the PVST with the difference that instead of word parts, prepositions must be invoked.

## 9. Discussion and future research

Having six equivalent tests assessing young learners' knowledge of PVs makes it possible to begin the piloting process of the tests. As in my doctoral dissertation (Thékes, 2016) the complex vocabulary was piloted, the same procedure will be taken in case of the PHAVE Test for YLs. The fact that each PV is assessed in three tests out of the six in the test battery, it will be made possible to compare the knowledge of PVs in different modalities. It will be empirically investigated whether form recall is the most difficult modality as Laufer et al. (2004) surmise.

First a paper-and-pencil pilot study will be conducted with the participation of around 120 6th graders, each sitting for three of the six tests. Test versions will be assigned randomly. Students' achievements will be examined and besides that, item level analysis of the test results will be conducted with classical statistical procedures and with the Rasch-model. Item-total correlation values will also be investigated to explore how the items behave in the tests. Checking these values, a clear picture will be revealed whether any change needs to be made on the test concerning any of the items.

Once the statistical analysis is finished, the tests will be uploaded onto an online platform called eDia (Molnár, 2014) and a large-scale online assessment will be done with 350 students. Besides the quantitative statistical procedures, think-aloud protocols will also be performed with twenty participating students so that more profound data will be revealed. The PHAVE Test for YLs will be a valid tool assessing Hungarian 6th graders; however, reproduction studies will also be made possible to be run at an international level.

Having developed a diagnostic test assessing the knowledge of phrasal verbs, I will provide facilitation for teachers' work. Teachers will have a new instrument at their disposal with which they can identify individual differences among students with respect to the acquisition of phrasal verbs. Teachers will, thus be shown the right road to applying more efficient methods to teach PVs.

## References

- Alderson, J. C. (2005): Diagnosing foreign language proficiency: The interface between learning and assessment. London: Continuum.
- Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E. (1999): Longman grammar of spoken and written English. Harlow: Longman.
- Boers, F., Demecheleer, M., Coxhead, A., and Webb, S. (2014): Gauging the effectiveness of exercises on verb-noun collocations. *Language Teaching Research*, 18(1), 50-70.
- Cameron, L. (2004): Teaching languages to young learners. Cambridge: Cambridge University Press.
- Csizér, K., and Dörnyei, Z. (2005): Language learners' motivational profiles and their motivated learning behaviour. *Language Learning*, 55(4), 613-659.
- Davies, M. (2010): The corpus of contemporary American English. Provo, Utah: Brigham Young University Press.
- Dunn, L. M., and Dunn, D. M. (2007): The Peabody Picture Vocabulary Test (PPVT-IV): Retrieved on 18th August, 2014 from <http://www.pearsonclinical.com/language/products/100000501/peabody-picture-vocabulary-test-fourth-edition-ppv-4.html>
- Fazekas, M. (2009): Felmérés a 2008-ban Magyarországon érettségizettek idegennyelv-tudásáról. Budapest: OKM (Manuscript).
- Gardner, D., and Davies, M. (2007): Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly*, 41, 339–359.
- Garnier, M. and Schmitt, N. (2015): The PHaVE List: A pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research*, 19(6) 645 –666.
- Harmer, J. (2012): Essential teacher knowledge. London: Pearson.
- Jimenez Catalan, R. M., and Terrazas Gallego, M. (2008): The receptive vocabulary of English foreign language YLs. *Journal of English Studies*, 5(1), 173–191.
- Kilgarriff, A. (1997): Putting frequencies in the dictionary. *International Journal of Lexicography*, 10(2), 135–155.
- Laufer, B. (2001): Quantitative evaluation of vocabulary: what it is good for and how it can be done. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara, and K. O'Loughlin (Eds.), *Experimenting with Uncertainty* (241-250). Cambridge: Cambridge University Press.
- Laufer, B., and Nation, I. S. P. (1995): A vocabulary size test of controlled productive ability. *Language Testing*, 16(1), 33-51.
- Laufer, B., and Nation, I. S. P. (2001): Passive vocabulary size and speed of recognition. *EUROSLA Yearbook* 1, 7–28.
- Laufer, B., Elder, C., Hill, K., and Congdon, P. (2004): Size and strength: do we need both to measure vocabulary knowledge? *Language Testing*, 21(2), 202–226.
- Liu, D. (2011): The most frequently used English phrasal verbs in American and British English: A multicorpus examination. *TESOL Quarterly*, 45, 661–688.
- Molnár, Gy. (2013): Számítógépes játékon alapuló képességfejlesztés: egy pilot vizsgálat eredményei. [Computer game-based skill development. Results of a pilot study]. *Iskolakultúra*, 21(4), 3-12.
- Nation, I. S. P. (2001): Learning vocabulary in another language. Cambridge: Cambridge University Press.

- Nation, I. S. P., and Beglar, D. (2007): A vocabulary size test. *The Language Teacher*, 31(7), 9–13.
- Nation, I. S. P., and Macalister, J. (2010): *Language curriculum design*. New York: Routledge.
- Nikolov, M. (1999a): Hungarian children's learning strategies. *Strani Jezici Zagreb*, 28(3-4), 225-233.
- Nikolov, M. (2007): A magyarországi nyelvoktatás-fejlesztési politika: Nyelvoktatásunk a nemzetközi trendek tükrében. [Language education development policy in Hungary] In I. Vágó (Ed.), *Fókuszban a nyelvtanulás [Language learning in focus]* (43–72). Budapest: Oktatókutatató és Fejlesztő Intézet.
- Nikolov, M. (2008): „Az általános iskola, az módszertan!” Alsó tagozatos angolórák empirikus vizsgálata. [„Primary school is real methodology”. Empirical investigation of English classes in lower primary school]. *Modern Nyelvoktatás*, 10(1-2), 3-19.
- Nikolov, M. (2011): Minőségi nyelvoktatás – a nyelvek európai évében. [Quality language education – in the European year of languages]. *Iskolakultúra*, 11(8), 3-12.
- Nikolov, M., and Mihaljevic Djigunovic (2006): Recent research on age, second language acquisition, and early foreign language learning. *Annual Review of Applied Linguistics*, 26, 234-260.
- Nikolov, M., and Szabó, G. (2011): Az angol nyelvtudás diagnosztikus mérésének és fejlesztésének lehetőségei az általános iskola 1-6. évfolyamán [Possibilities of developing English diagnostic tests for years 1-6 in the primary school]. In B. Csapó and A. Zsolnai (Eds.), *A kognitív és affektív fejlődés diagnosztikus mérése az iskola kezdő szakaszában (13-40)*. Budapest: Nemzeti Tankönyvkiadó.
- Orosz, A. (2009): The growth of young learners' English vocabulary size. In M. Nikolov (Ed.), *Early learning of modern foreign languages. Processes and outcomes (181-195)*. Bristol: Multilingual matters.
- Qian, D. (2002): Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning* 52(3), 513–536.
- Schmitt, N. (2008): Instructed Second Language Vocabulary Learning. *Language Teaching Research*, 12(3), 329-363.
- Schmitt, N., Jiang, X., and Grabe, W. (2011): The percentage of words known in a text and reading comprehension. *Modern Language Journal*, 95(1), 26-43.
- Schmitt, N., Schmitt, D., and Clapham, C. (2001): Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language Testing* 18(1), 55–88.
- Singleton, D. (1999): *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.
- Thékes, I. (2015a): An integrated diagnostic vocabulary test for young EFL learners. *Journal of Language Teaching and Linguistics*, 6(2), 189-209.
- Thékes, I. (2015b): Angol, mint idegennyelvi szótudás mérése egy diagnosztikus komplex online teszttel fiatal nyelvtanulók körében. *Iskolakultúra*, 25(12), 28-56.
- Thékes, I. (2016): *Assessing young Hungarian EFL learners' vocabulary and learning strategies*. Unpublished doctoral dissertation. Szeged: University of Szeged.
- Thornbury, S. (2004): Big words, small grammar. *ELT Professional*, 31(1), 10-11.
- Vidákovich T., Vígh T., Sominé Hrebik O., and Thékes I. (2013): Az angol és német nyelvi szókinccs online diagnosztikus tesztelése a 6. évfolyamon. [Diagnostic assessment of English and German as a foreign language vocabulary amongst 6th graders]. *Iskolakultúra*, 23(11), 117-131.
- West, M. (1953): *A general service list of English words*. London: Longman.

Xue, G., and Nation, I. S. P. (1984): A university word list. *Language Learning and Communication*, 3(2), 215-229.

Zareva, A. (2007): Structure of the L2 mental lexicon: How does it compare to native speakers' lexical organization? *Second Language Research*, 23(2), 123-153.

## Author

**István Thékes**, Gaál Ferenc University, Szeged, Hungary. E-mail: thekes.istvan@gff-szeged.hu

## Appendix 1. One task in the COPHAVE battery

### Task 1

Instruction: Choose the phrasal verb that matches the meaning of the verb in bold. See example

Example: He is sitting now but he will **be on his feet** in a second.

- A) bring over
- B) get off
- C) get down
- D) stand up

Correct answer: D)

1) Uncle Jim **will go to his house and give us** our lawn-mower.

- A) bring back
- B) get up
- C) get down
- D) go ahead

2) Prices usually **increase** year by year

- A) come back
- B) get up
- C) go up
- D) move in

3) My mom has travelled to England but will **return** soon.

- A) bring back
- B) come back
- C) give in
- D) go ahead

4) The DJ has asked everybody to **move onto the dancefloor**.

- A) bring back
- B) get down
- C) get up
- D) go up

5) The football player is on the field; he has to **lift his body** and play on.

- A) get up
- B) go ahead
- C) go off
- D) move in



6) Her grandpa has been smoking. He will need to **stop** doing so.

- A) come back
- B) give up
- C) go up
- D) move in

7) - Can I have some of your hamburger?

- Sure, **help yourself**.

- A) get up
- B) give in
- C) go ahead
- D) go up

8) The criminals are threatening the city council that a bomb will **explode** outside their office center.

- A) get down
- B) give in
- C) go ahead
- D) go off

9) The family cannot wait to start to **live in their new house**.

- A) come back
- B) get up
- C) go off
- D) move in

#### Task 2

Instruction: Choose the correct preposition after the verb. See example.

Example: He is sitting now but he will stand.....soon.

- A) in
- B) on
- C) over
- D) up

Correct answer: A)

1) I need new information. I have to find.....what they are up to.

- A) back
- B) on
- C) out
- D) over

2) I have been waiting here. I hope I will get.....within an hour

- A) back
- B) down
- C) in
- D) over

3) I tried to convince the president last night. I hope he will give.....and we can have his signature.

- A) back
- B) down
- C) in

D) over

4) The child was acting badly. The teacher asked him to go.....

- A) in
- B) into
- C) out
- D) over

5) If you fix a meeting on Facebook messenger, you have to show .....and talk to your friend.

- A) in
- B) on
- C) up
- D) through

6) If your plastic mineral water bottle is empty, you have to throw it.....not to pollute the environment.

- A) across
- B) down
- C) in
- D) out

7) Try to remember the cell phone number of your friend. Write it.....not to forget.

- A) down
- B) in
- C) into
- D) over

8) Try to walk.....quietly so the baby will stay asleep.

- A) back
- B) in
- C) out
- D) over

9) When you ride your bike, never look.....so that you will always focus on the road.

- A) back
- B) in
- C) over
- D) through

### Task 3

Instruction: Write the proper preposition into the gaps. See example.

Example.

When his Iphone rings, He will stand....and walk to pick it up.

Correct answer: up

1)The airplane will take.....in a minute and head to Beijing.

2) My Blackberry alarm app will wake me.....at exactly 6 o'clock.

3) Sarah will put.....he new Hello Kitty denim vest this morning.

- 4) When my children grow....., they will not watch Wild Kratts cartoons.
- 5) The monkey has climbed the tree in the zoo not he has to come.....from above.
- 6) Corn flakes is always great to start.....the day.
- 7) Come.....boys, we can win this match.
- 8) Vincent talks too much in class, sometimes he needs to shut.....so everyone can focus on the lesson.
- 9) We are going to Venice on holiday. On the first day, we set.....in the morning.