Doctoral Program of Info-Communication Technologies
Doctoral School of Education
University of Szeged

Assessing Young Hungarian EFL Learners' Vocabulary and Learning
Strategies

PhD dissertation theses

Thékes István
Consultant: Nikolov Marianne, D.Sc.

Szeged
2016

Introduction

The 1990s saw an increasing number of studies focusing on foreign language (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; Schmitt, 1997; Thékes, 2015a; Thékes, 2015b; Vígh, S. Hrebik, Thékes & Vidákovich, 2015) the FL mental lexicon (Singleton, 1999; Zareva, 2007), corpus studies (Kilgarriff, 1997; Nation & Macalister, 2010; Thékes, 2014a) and vocabulary learning strategies (VLS) (Nation, 1990; Schmitt, 2008; Thékes, 2016). It has also been affirmed that vocabulary knowledge is a good predictor of reading comprehension (Nassaji, 2003). It has been reported that receptive vocabulary knowledge predicts productive vocabulary knowledge (Thékes 2015a). With the emergence of the lexical approach (Lewis, 1993) in language teaching, a new avenue was opened for vocabulary research. The learning and teaching of vocabulary is a popular research area in the FL learning literature. These two processes are in the center of attention of both scholars and teachers.

Purpose of research

Most of the validated diagnostic vocabulary tests had been originally used as paper-and-pencil tests and there is a lack of vocabulary measures carried out in an online environment. Albeit the paper-and-pencil vocabulary tests had been computerized and are available on Tom Cobb's website at www.lextutor.ca, hardly any study has been published on the assessment of vocabulary executed with an online instrument except for Vidákovich, Vígh, S. Hrebik & Thékes, 2013). There was a need to develop an online English as a FL vocabulary test assessing YLs' receptive and productive word knowledge. Besides developing and validating this vocabulary test, there was also an intention to see correlations between YLs' word knowledge and word learning strategies. A questionnaire was developed that looked into YLs' foreign language VLS in an online environment is taken by the students after they sit for the online vocabulary test. In order to procure a profound picture of YLs' EFL vocabulary and VLS, the results had to be triangulated. Think-aloud protocols, interviews were conducted, plus teachers were requested to assume YLs' EFL vocabulary and VLS. The vocabulary test and the VLS questionnaire were piloted and validated on paper and pencil.

Literature

It is postulated by Read (2000, p. 32) that there are two contrasting perspectives of vocabulary assessment. One viewpoint is that vocabulary items can be tested as an independent semantic field independent of context. The other view is that lexis must always be measured in context. This issue should be the concern of test-givers. The issues emerging from language testing research need to be looked at from four different angles (Lehmann, 2009; Nation, 2013; Milton & Fitzpatrick, 2014). Four major questions are proposed by Nation (2013) that need to be addressed: (1) why to test vocabulary? (2) what words to test? (3) what aspects of word knowledge to test? (4) how to test the various aspects of word knowledge? Before presenting the data-gathering instrument, and the findings of studies assessing the word knowledge of young learners (YLs), I will elaborate on the characteristic traits and principles of diagnostic testing of FL in the context of YLs. Nikolov and Szabó (2011). These principles had been grounded the study by Alderson (2005) and McKay (2006, p. 36) who states that diagnostic assessment of YLs' FL proficiency and word knowledge is important because a lot can be implied from them in classroom practices. I will attempt to synthesize six of these traits which, I believe, are the most relevant from the perspective of computerized vocabulary assessment of YLs: 1) the purpose of diagnostic tests is to identify the strengths and weaknesses of learners; 2) 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; 3) diagnostic tests are low-stakes tests or consequences are of irrelevant weight so optimal achievement is not hindered by anxiety or any other affective factor; 4) diagnostic tests are more likely to focus on 'lower-level' linguistic abilities than on 'higher-level' abilities; 5) diagnostic testing is probably made more efficient by using a computerized platform.

Pilot studies
Pilot study of the vocabulary test

The purpose of the pilot study was three-fold:
1) to develop a complex diagnostic vocabulary test for YLs of English as a FL
2) to find out how the different items work by means of item-analysis; thus to validate it
3) to implement a test that will be used online ultimately

When developing the tasks for the vocabulary test I took Laufer et al.'s (2004) task taxonomy into consideration who discerned four degrees of knowledge of meaning 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 dissertation Schmitt's (2014) terminology is used.

To select the words into the test battery, 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 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 BNC; 2) occurrence of the words in course-books used by 6th graders; 3) personal judgement on the assumed difficulty of the word. Selection of the lexical items for the tasks was completed by choosing words from all three frequency categories. In all tasks the majority four or five words belonged to Category 1 and at least one word represented Category 3, which means that Category 2 included three or four items. With this system, it was guaranteed that words form all the possible categories were assessed.

The students taking the test were 6th graders (n=103) in four Hungarian primary schools in a convenience sample in Szeged and Mezőtúr. Careful selection took place in terms of number of English lessons per week. Only students in classes of general curriculum were selected. This means that learners had three English lessons a week in the school-year when data collection was carried out and they had been learning English since 4th grade.

A diagnostic vocabulary test was designed to assess learners' word knowledge. Most of the diagnostic vocabulary tests measure one dimension of vocabulary (Nation, 1990). They either tap into receptive or productive word knowledge. The diagnostic instrument consisted of seven different tasks.

Besides taking corpus-based data into account, recommendations in the Hungarian curriculum and Nikolov (2011) were also considered in terms of grouping words based on topics and involving them in the list.

Six tasks (Task 1-Task 6) of this complex vocabulary test were intended to assess breadth of vocabulary since most vocabulary tests (Thékes, 2015b) assess this domain. One task (Task 7) was intended to assess depth of vocabulary. The required word knowledge for solving task was receptive in the first five tasks and in Task 6 and 7 productive word knowledge was required. The VKS was implemented in Task 7. Moreover, I reckoned that it would have carried a heavy cognitive load for 6th graders if I had tested depth in more than one task.

The vocabulary test battery was administered in four schools in seven classes in November 2013. Language classes in Hungary are usually divided into two groups with two teachers working

with the separate groups simultaneously. However, test taking took place in whole classes in order to save time.

The reliability of the test battery proved to be acceptable (Cronbach's Alpha= .82). The most reliable task proved to be a listening task, Task 2 (Alpha=.87) followed by a reading task, Task 4 (Alpha=.82). The least reliable task was Task 7 (Alpha=.48), the one assessing vocabulary depth. Task 7 included malfunctioning items with low item-total correlation, thus this reliability value is not unexpected.

High and significant correlations were revealed. The two listening tasks (Task 1 and 2) gave an indication of a robust statistical relationship with each other, whereas the two reading tasks (Task 4, and 5) had a much weaker, but still significant, correlation. These significant correlations prove that tasks addressing similar language skills are very similar to each other; consequently, they measure the same construct. If the items were shifted from Task 2 to Task 1 and conversely, approximately identical results would be achieved by the students. As far as correlations amongst the seven tasks are concerned, high correlations can be pointed out between some task types. Results indicate that other high correlations (.85) were noticed between Task 3 and Task 5, two reading tasks and between Task 1 and Task 4, a listening task in meaning recognition modality and a reading task in meaning recognition modality. The strong relationship between the two listening tasks can be explained by the similar nature of the two tasks. Task 3 and Task 5 were also similar in the sense that meaning recognition was expected from the students through reading.

It was also necessary to study the functioning of the depth of vocabulary test (Task 7) since it indicated significantly negative correlations with all the other task types except Task 4 in which students were expected to match written words with pictures. Negative correlations mean that the items in Task 7 do not function adequately and their responses do not vary in line with those for the other tasks.

After examining the descriptive statistics of the test, especially the item-analysis, a decision was made to replace some of the items in the test. Since the goal of the test development, pilot study and the item-analysis was to create an instrument adapted for online use, items were carefully analyzed. Items with low standard deviations and low item-total correlation, under .20 (Field, 2005), were omitted from the test and these items were then replaced with new words. It was ascertained that no cognates would be used again since the cognate 'hamburger' elicited a ceiling effect and the new items had the same category as the item it was replaced for. It was also an obvious outcome of the pilot study that Task 7 did not function as acceptably. Its reliability was low and the negative correlations of this task with the rest of the other six tasks led me to the conclusion that leaving Task 7 in the online instrument would risk the validity of the test. Hence Task 7 was omitted from the online version. This also meant that the instrument lost nine items; however, the remaining six tasks still comprised 68 components and I judged this amount would be sufficient in the online test. Apart from getting rid of one task, the format of the instrument was not modified and the same six tasks were applied for further data collection.

Pilot study of the questionnaire investigating YLs' vocabulary learning strategies

Having investigated the instruments assessing VLS with special regard to those of YLs, a decision was made to consider Stoffer's (1995), Schmitt's (1997) and Pavičič's (2008) questionnaire items adapted from Oxford's SILL (1991) for a large item pool. The reason for this was that these instruments had been either used or adjusted for YLs VLS were concerned. The pool also comprised items that were considered worthy of being a component of a questionnaire assessing Hungarian YLs' VLS. The items from all of the selected questionnaires were considered for inclusion in the new instrument.

Five factors were selected to be the composing parts of the questionnaire: cognitive strategies, strategies involving memory, metacognitive strategies, strategies involving determination and social. Once the five dimensions had been settled upon, the questionnaire statements were carefully phrased with a view to the specific Hungarian learning environment and circumstances. Two experts were consulted during the process of questionnaire development. All the questionnaire items were thoroughly thought over with special respect to their wording in Hungarian, the NL of the YL participants, so that they would reflect the construct.

Following the selection of the questionnaire items it was also decided that the data would be assembled on a 4-value frequency scale: 'never', 'once a month', 'once a week', 'always.' The decision was made with the intention of forcing to students not to opt for a neutral answer. Since the foundations

of my questionnaire were laid on Oxford's (1991) and Schmitt's (2008) data collection instruments, their way of data collection ought not to be left out of consideration.

The children took the questionnaire seriously and filled in it. When the completed questionnaires were collected, I asked each student to write down strategies that they use to learn words on their own. The lists of strategies of all the students were later considered for inclusion in the final questionnaire. This resulted in new items being involved in the modified instrument used in the large-sample assessment.

The pilot study was carried out with the participation of 86 Hungarian 6th graders in primary schools in Budapest, Mezőtúr and Szeged in February 2014. All the students had studied English from their 4th grade (age 10) in three lessons a week.

The reliability of the questionnaire was fairly high (Cronbach's Alpha = .91). It was also concluded that some of the items had 0 standard deviation. Every student indicated 'always' at the statement 'I use a vocabulary list to learn word'. Since this type of item provides the research with no information from the perspective of educational science, it was decided that items having zero standard deviation would not be used in the final questionnaire. The correctness of the decision on adding the item 'My parents check if I have learned the new words by asking me' was justified, since participants reported high frequency of this activity (M=3.09). Dictionary use also appeared to be a frequent activity used by students. Both the item 'I look up the new word in an English-Hungarian dictionary' and 'I look up the meaning of the word in an electronic dictionary' had high frequencies (3.09 and 3.03, respectively) as it had been previously assumed Hungarian learners have a tendency of using dictionaries for the purpose of learning words. The activities 'I infer the meaning of the new word from context when reading' and 'I infer the meaning of the new words from spoken English', also turned out to be often used by learners (M=2.86 and M=2.84, respectively).

Following the investigation of descriptive statistical data and the frequencies of the different items, item-analysis was carried out by means of scrutinizing corrected item-total correlations. On a sample of 103 students, the reliability and the validity of the items with values under .20 are endangered on account of the fact that these items work differently from the given construct (Field, 2005). A decision was made item by item as to which items that fell under the value of .20 within its own factor would be omitted from the questionnaire and those that fell near this value would further be examined.

Even though the theoretical model was not entirely justified by the exploratory factor analysis, it was decided that the original model would be kept and in the large-sample study a confirmatory factor analysis would be conducted. The reason for this decision was two-fold: 1) the new clusters following the exploratory factor analysis did not form any new interpretable dimensions; 2) the clusters created on the basis of the factor loadings did indicate some confirmation of the theoretical model.

The new instrument was developed based on the new factors and it was uploaded onto the eDia platform in order it would be used on a large sample. Having investigated the descript tive statistics of the questionnaire results and having gained an insight into the items, factor analysis was conducted to check whether the five factors reflect the original conceptualization. After the factor-analysis had been run, it turned out that ten factors existed on the basis of the results. The KMO-index was .72 which was an indication that the strength of the correlation among the five dimensions makes it moderately adequate for factor analysis.

Based on the item-correlation indices the items of low values were removed from further investigations and three new items were added to the remaining ones which were then designated under the pertaining factors. The new items were the ones mentioned by the students participating in the pilot studies.

The large-sample studies
Research questions

Having piloted the vocabulary test and the questionnaire, a large-sample assessment was conducted to map YLs' EFL word knowledge and vocabulary learning strategy use, and ultimately to reveal correlations. The following research questions were phrased.

1) How does the YLs' performance on the vocabulary test explain EFL vocabulary size?
2) How can conclusions be drawn from students' achievements as regards the way items function on the vocabulary test?
3) From a criterion-referenced testing perspective, how do students know the most frequent English words?
4) How do the relationships amongst tasks of different modalities provide an insight into the construct of YLs' EFL word knowledge?
5) How do the high-achieveing students perform on the productive task of the vocabulary test?)
6) How do teachers estimate the vocabulary size of 6th graders?
7) How reliably does the quantitative instrument assess the self-report word study strategies amongst YLs?
8) Which strategies are the most frequently used ones?
9) Which strategies are used less frequently?
10) How can implications be drawn from the correlations of the the factors of the word strategy questionnaire correlate with one another?
11) What VLS do teachers assume students use?
12) How do factors of word study strategy use and other background variables explain vocabulary size?
13) How do the correlations of different tasks of the vocabulary test with word study strategy use explain vocabulary size?

Participants

The sample was selected by the coordinators of the Institute of Educational Science. The Institute filed a request to schools in Hungary and twelve schools agreed to involve their students in the research. Participants were 282 Hungarian 6th graders. Sampling was nonrepresentative.

Instruments

The two major instruments (the vocabulary test and the VLS questionnaire) were piloted in November 2013 with the participation of 103 6th graders. The test development was grounded on corpora (word frequency ranks, curriculum and expert advice). The VLS questionnaire was developed on the basis of the literature and expert advice.

Seven data collection instruments were used in the study:
1) the online vocabulary test comprising six tasks
2) think-aloud protocols elicited during taking the vocabulary test to map the students' thought processes
3) a paper-and-pencil questionnaire for teachers related to the vocabulary test to gain insight into teachers' assumptions on students' vocabulary size
4) the online VLS questionnaire
5) the paper-and-pencil questionnaire of teachers' beliefs
6) interviews with students reporting their VLS use

7) a background questionnaire inquiring into learners' EFL motivation, school grades, number of EFL lessons a week, etc.

Procedures

The volunteering schools were given a passcode to be able to log into the eDia platform where the vocabulary test, the vocabulary learning questionnaire and the background questionnaire could be accessed. Data were gathered in November 2014 and data processing was performed with the use of the SPSS 17 software.   The platform called eDia is undergoing constant development (Molnár, 2013) and is adequate for efficient data gathering on a large sample. The sound files were also attached to the first two tasks of the vocabulary test. My voice, the researcher's, was recorded reading out the items. Every task contained a sample task that was presented to the students before they went about taking the test. Taking the vocabulary test took approximately fifteen minutes and filling in the online questionnaire also took this amount of time. Students sat down in front of the screen with head-sets over their ears so that they could hear the voice file of the first two tasks.

Results of the vocabulary test

In all the six tasks, except for Task 3, there were nine items plus one item was an exemplary item and one was a distractor; as a result test-takers had to know nine items. In Task 3, students had to know nine items and there were nine distractors and six exemplary items. In every task the maximum points were nine. This meant that the maximum points in the whole test were 54 points. Reliability of the test proved to be acceptable (Cronbach's Alpha = .869).

As regards Task 1, there was significant difference between this task and Task 2 and Task 4, and Task 5, and Task 6. As far as Task 2 is concerned, the scores on this task differed significantly from those in Task 1, as mentioned before, and Task 3. As for Task 3, there was significant difference in scores between this task and Task 2 and Task 4, and Task 5, and Task 6. There was no significant difference among the scores of Task 4, Task 5, and Task 6.

Contrary to the paper-and-pencil pilot study, on the online test with a larger sample size, participants had the best achievement on Task 1. In the pilot study, Task 3 proved to be the task where students had the highest achievement. Nonetheless, Task 1 and Task 3 proved to be the easiest tasks during both test procedures. Both tasks are completed in meaning recognition modality which is assumed to be the easiest in the hierarchy of modalities (Laufer at el., 2004; Schmitt, 2008).

It ought to be highlighted that the item-total correlation values of all items except for 'icecream' proved to be acceptable, i.e., above the .20 limit (Field, 2005). Even though some items were in the vicinity of this critical value (e.g., 'lion', 'sausage', 'cleaning' and 'busdriver') the instrument does not appear to suffer from low item-total correlation values, thus it can be claimed that the entire instrument yields valid results. It is a remarkable fact as well that items with low item-correlation values have an even distribution across tasks. No task has more than one item that works inconsistently with the average functioning of the other items. This might provide evidence for the fact that the tasks requiring different task-solving modalities have equal strength and assess the same construct.

As for the correlations among tasks, the tasks assessing listening, Task 1 and Task 2 are in a significant correlation (r=.50, p<.01), which means that whether form or meaning recignition are the expectation, the two tasks assess the same construct. The two tasks demanding reading skills, Task 4 and Task 5 also significantly correlate with each other in a slightly weaker relationship (r=.36; p<.05). Two tasks that expected the students to match pictures with words (Task 1 and Task 4) also significantly correlated (r=.33, p<.01). However, two tasks in which students had to match definitions with words (Task 2 and Task 5) did not significantly correlate. It might be problematic that two tasks that measure the same construct within one instrument do not correlate with each other. This outcome seems to justify the supposition that task assessing word knowledge through listening and reading measure two different abilities (Vidákovich et al., 2013).

Having analyzed the results of the items with the means of classical test theory, the applicability of the tools of modern test theory was also considered. The Rasch-model was employed to gain a deeper

insight into the reliability of the test and the difficulty of each item. The Conquest program was used to conduct the Rasch-analysis. The value used in modern test theory, EAP/PV, of .912 yielded evidence of high reliability. As 54 items were assessed in the test, the same number of items are scaled by logits. Based on the model, the assertion can be made that the test has a relatively normal distribution. Most of the items are in or near logit 0, which indicates a normal distribution. As regards easiness and difficulty of items, easy and difficult items are represented almost in an equal ratio, so the test differentiates properly.

Results of the words study strategy questionnaire

A profound insight could be gained into YL's VLS use by examining dimensions established within the theoretical framework. It was analyzed which of the five strategy dimensions (cognitive, memory, metacognitive, social, determination) outlined by Schmitt (1997) was reported to be used the most frequently and the reliability of the factors within questionnaire was also examined. The reliability of the questionnaire was acceptable (Cronbach's Alpha = .81)

Students reported using memory strategies the most often. This finding debunks the outcomes of the research conducted by Doró and Habók's (2013) who asserted that metacognitive strategies were most often used by YLs. However, they also found that memory, cognitive and social strategies were almost equally often applied by YLs. In this research, it stands out that social strategies are less frequently used than any other strategy. The plausible reason for memory strategies being the most frequently used ones is that YLs learn FL words in chunks and memorizing these chunks is a crucial way of acquiring them (Wray, 2002). My finding is, in contrast to Doró and Habók (2013), confirmed by Lan (2006) who stipulated that memory strategies were most often used by Taiwanese YLs (age=11).

In order to triangulate the data, 18 teachers of English teaching in 6th grade were requested to approximate how YLs learn English vocabulary. Only teachers teaching YLs were asked intentionally so that relevant data would be procured in this field. Data were gathered in June 2015. The ten most frequent and the five least frequent strategies reported by the YLs on the online questionnaire were listed and this list was given to the teachers who saw the question: 'How often do you believe students use these strategies?

Comparing the mean values of the teachers' and the students' questionnaire it can be concluded that what teachers believe is not distant from what students reported. Those strategies that were reported being used not so frequently by students are exactly the ones that teachers believe students hardly ever use. The fact that teachers' views are in parallel to what students do in terms of FL word learning strategy use is a positive fact since efficient teaching takes place when teachers know what their students actually do (Borg, 2003, p. 82). Similarily to the least often used strategies, teachers reported approximately analogous data to those of the students as regards the most often used strategies. Teachers are very well aware of the fact that students repeat the words to themselves so as to learn them and that students very often seek to remember the Hungarian equivalent of the English words.

Validating the questionnaire with confirmatory factor analysis

Confirmatory factor analysis (CFA) was applied to examine the underlying measurement model of vocabulary learning strategies. Different fit indices, the Tucker-Lewis Index (TLI), the comparative fit index (CFI) and the root mean square error of approximation (RMSEA), were computed to provide necessary information in determining model fit. Weighted Least Square and Mean- and Variance-adjusted (WLSMV) estimation was used (Muthén & Muthén, 2010).

First a five-dimensional model was investigated with the five factors. The five-dimensional model did not show a good model fit according to the CFI and TLI indices. The inconsistence in the functioning of the items belonging to the social strategies factor made it suspicious that the social factor ought to be discarded. The CFA with the five-dimensional model (with the social strategies being part of it) gave evidence to this suspicion. Due to the low covariance values of the social factor in the matrix, its low factor-loadings and the improvable CFI and RMSEA values, a decision was made to discard the factor and to run a four-dimensional model. After conducting the CFA with the four-dimensional model I found that it proved to have better model fit.

Main findings of the research

RQ 1) It was found based on the t-tests that Task 1, the listening task of meaning recognition modality, proved to be significantly the easiest (M=6.39) and significantly the most difficult task was Task 4, a reading task of meaning recognition modality (M=2.75). It was asserted during data analysis that a task of form recall (Task 6), a supposedly difficult task, proved to be significantly significantly easier (M=3.38) based on the t-tests.

RQ 2) The item-total correlation values were evaluated. This calculation is performed to check if any item is inconsistent with the other items. The minimum of this item-total correlation value is .20. None of the items, except for 'icecream' (.01) fell below this value. In case a test is under development, it is suggested that the items below .20 should be discarded. In this case there was no possibility to replace 'icecream'; however in further research a new item will be used in Task 6. Some very low values are unveiled. The item, for example, the most learners knew, 'clean' had a value of only .26. 'Lion' was also near the critical limit with an item-total correlation value of .27.

RQ 3) Students' overall mean knowledge of Category 1 words was 48.34 (SD=43.73). It was also asserted that some students (see the ones in the high achieving tercile) knew over ninety percent of the Category 1 words. It is obvious that the mastery of Category 1 words is inevitable in the process of learning less frequent words and of progressing to higher ability levels in EFL. The critical limit of proceeding to next stage of learning was set at the knowledge of 80 % of Category 1 words based on Nagy (2004). It was found that out of the 282 participants 108 (38.29%) knew at least 80 % of Category 1 words. It was stipulated that the rest of them (61.71%) ought to endeavor into learning the most frequent English words until they reach the critical value, i.e., 80 % of Category 1 words.

RQ 4) The six tasks indicated significant correlations with one another with the exception of Task 5 and Task 6. Task 5 had a weak relationship with Task 2 (r=.01) and a strong relationship but no significant correlation with Task 1 and Task 3 (r=.06 and r=.06, respectively). Task 6 had a weak relationship with Task 2 (r=.11) and no significant correlation with Task 3 and Task 4. Task 1 and Task 2, the two listening tasks yielded an indication of a modest relationship with a significant correlation (r=.50, p<.01), meaning that no matter whether the modality was meaning recognition or form recognition, the two tasks measured a similar construct. Task 4 and Task 5 also correlated significantly with a weaker relationship (r=.36, p<.05).

RQ 5) It was expected that the productive task in form recall modality would be the most difficult task and as such it would be a major differentiating factor among the participants of different word knowledge. Task 6 did not prove to be the most challenging task. However, I intended to know how students in the high achieving tercile performed on this particular task to gain better insight into the organization of their vocabulary. Students in the high achieving tercile had a mean of 4.78 on the productive task, which means that they reached nearly 50% on this task. It is a low value compared to the number of points they reached on Task 1, Task 2 and Task 3.

RQ 6) It can be asserted that 18 teachers of English of 6th graders generally overestimated the word knowledge of students, following the comparison of the results of what teachers assumed and what students achieved. Besides examining the descriptive statistics of teachers' assumptions and students' test scores, the t-test also confirmed the existing over-estimation of students' word knowledge. It was highlighted that students scored a mean of 6.39 on the online test whereas teachers expected them to score over 7.38. The assumed scores were characterized by an over-estimation on part of the teachers.

RQ 7 and RQ 8) These RQs enquired into the ratio of occurrence of VLS. The memory strategy 'I repeat the word to myself' is the most frequently used strategy followed by the strategy 'I try to remember the Hungarian equivalent of the new English words.' The strategies reported by students 'I use a new word in speaking so as to remember it' and 'I listen to English music in order to learn new words' are also

very frequently used by learners as it was reported by them. In order to gain information regarding the distribution of the frequency of strategy use of YLs, it was also investigated which strategies students reported using less frequently. Asking classmates what the new word means is a strikingly infrequent strategy. Making English-Hungarian word cards goes out of fashion in terms of strategy use as it was reported to be used very rarely. The strategy 'I use Skype to learn English words' is also an opportunity not exploited by the YLs.

RQ 9) As for correlations across the factors in the questionnaire, it was found that social strategies showed very weak relationship and no significant correlations with the rest of the factors. This is an indication that social strategies are not part of the internal structure in the questionnaire. The insignificant and low correlational values pointed to the fact that the items of social strategies function discrepantly in relation to the other items in the instrument. This might give rise to considering discarding this factor from the questionnaire. In case of all the other four factors a significant correlation was found among one another. It was also highlighted that there was not a strong relationship among them with r-value .49 being the highest in the correlational matrix.

RQ 10) For the sake of triangulating data with respect to students' vocabulary strategy use, not only quantitative data gathering and interviews were conducted but teachers were also requested to express their beliefs on the vocabulary strategy use of YLs. It was revealed that teachers believe that students use the listed strategies most frequently: 1) repeating the new word to themselves, 2) using Skype to learn English words, 3) trying to remember the Hungarian equivalent of the English words, 4) using Facebook to learn words, and 5) learning words from their vocabulary.

RQ 11) The correlations between the vocabulary learning strategy use and the results of the vocabulary test were analyzed. Previous studies (Schmitt, 1997) pointed out that repetition and mechanical learning strategies were more popular than strategies requiring heavy cognitive load among YLs. However, popular strategies are not necessarily the ones that are needed for successful vocabulary learning; therefore, it was observed what strategies were used by students in the high achieving tercile to the greatest extent. Students in the high achieving tercile also frequently learn new words from their own vocabulary. The five most frequent strategies used by students in the high achieving tercile are 'I try to remember the Hungarian equivalent of the new English words' (M=3.47, SD=1.00), 'I repeat the word to myself' (M=3.23, SD=.94), 'I underline the important words (M=3.05, SD=1.21), 'I evaluate if I have really learned the word' (M=3.00, SD=1.28), and 'I learn new words from my own vocabulary' (M=3.00, SD=1.28).

RQ 12) The regression analysis indicated that the frequency of English classes is the strongest predictor of EFL word knowledge. Motivation was found to be the second strongest predictor of EFL word knowledge However, school grade in English was a background variable that was not a significant predictor of EFL word knowledge. As for the relationship between word study strategy use and the results scored on the vocabulary test, cognitive, memory and metacognitive strategies appear to have predictive strength; however, it seems that the use of determination and social strategies do not significantly predict EFL word knowledge.

Conclusion, limitations

The investigation of YLs' English as a foreign language vocabulary size was a major endeavor since an online data-gathering instruments had to be developed and created. Having conducted a pilot study with the two instruments, item-analysis and several statistical procedures were executed in order that a properly functioning test would be used on large sample for the sake of unveiling correlations and of gaining a deeper insight into the organization of vocabulary. Inevitably, there are a number of limitations that need to be noted in this study. First, the sample was not representative and a larger number of participants might have increased the reliability of both the vocabulary test and the self-report VLS questionnaire. This caveat of my research in some cases restricted the significance level of my results; therefore, the extension of the results to the population also encountered restrictions. On the other hand the sample size number of 288 is not as low as that in previous research so classroom implications might

be drawn from the results. The realization of the qualitative investigation involving think-aloud protocols and interviews plausibly were also limited on account of a small sample size. As   for classroom implications of the correlations between VLS and the vocabulary test, a few conclusions can be drawn. Since students in the high achieving tercile reported using such strategies frequently as 'I try to remember the Hungarian equivalent of the new English words', 'I repeat the word to myself', 'I underline the important words', 'I evaluate if I have really learned the word', 'I learn new words from my own vocabulary', and also 'play English video games', these strategies might be considered as efficient ones; however that does not mean that students achieve highly on the vocabulary test because they use these particular strategies. In addition, the sub-samples based on achievement use by and large the same types of strategies according to this study's findings. For the sake of the students' advancement in EFL word learning, strategy training is plausibly expedient. It is nowadays axiomatic that YLs use online media tools very often and their way of learning differs to a great extent from what their teachers applied when they were students. Thus, teachers themselves must be prepared and be aware of the fact that their students learn differently and use different strategies. Teachers' promoting online glossing, the use of online dictionaries and even conscious use of social media for the purpose of mastering words could be beneficial.

Literature

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.

Doró, K., & Habók, A. (2013). Language learning strategies in elementary school: The effect of age and gender in an EFL context. *Journal of Linguistics and Language Teaching 4*(2), 25-37.

Field, A. (2005). *Discovering Statistics Using SPSS*. 2nd ed. London: Sage.

Kilgarriff, A. (1997). Putting frequencies in the dictionary. *International Journal of Lexicography, 10*(2), 135–155.

Lan, R. L. (2006). *Language learning strategies profiles of EFL elementary school students in Taiwan.* PhD dissertation. Annapolis: University of Maryland.

Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: do we need both to measure vocabulary knowledge? *Language Testing, 21*(2), 202–226.

Lewis, M. (1993). *The lexical approach*. Cambridge: Cambridge University Press.

McKay, P. (2006). *Assessing young language learners*. Cambridge: Cambridge University Press.

Molnár, Gy. (2013). Számítógépes játékon alapuló képességfejlesztés: egy pilot vizsgálat eredményei. [Computer game-based kill development. Results of a pilot study]. *Iskolakultúra, 21*(4), 3-12.

Muthén, L.K., & Muthén, B.O. (2010). *Mplus User's Guide*. Los Angeles, CA: Muthén & Muthén.

Nagy, J. (2004). A szóolvasó készség fejlődésének kritériumorientált diagnosztikus feltérképezése. [The criterion-referenced, diagnostic mapping of the development of word-reading skills]. *Magyar Pedagógia, 104*(2), 123–142.

Nation, I.S.P. (1990). *Teaching and learning vocabulary*. Boston: Heinle and Heinle.

Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I.S.P. (2013). *Learning vocabulary in another language*. (Second edition). Cambridge: Cambridge University Press.

Nation, I.S.P., & Macalister, J. (2010). *Language Curriculum Design*. New York: Routledge.

Nikolov, M. (2011). Minőségi nyelvoktatás – a nyelvek európai évében. [Quality language education – in the European year of languages]. *Iskolakultúra, 11*(8), 3-12.

Oxford, R. (1991). *Language learning strategies: What every teacher should know*. New York: House/Harper and Row.

Pavičič, T. V. (2008). *Vocabulary learning strategies and foreign language acquisition*. Clevedon: Multilingual Matters.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 198-227). Cambridge: Cambridge University Press.

Schmitt, N. (2008). Instructed Second Language Vocabulary Learning. *Language Teaching Research 12*(3), 329-363.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning, 64*(4), 913–951.

Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.

Stoffer, J. (1995). University foreign language students' choice of vocabulary learning strategies as a related to individual difference variables. *Journal of Education of the University of Alabama, 4*, 122-153.

Thékes, I. (2014). Assessment of Young learners' English as a foreign language vocabulary knowledge: Results of a diagnostic test with the Rasch model. In J. Horvath & P. Medgyes (Eds.), *Studies in honour of Marianne Nikolov* (pp. 87-100). Pécs: Lingua Franca Csoport.

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. [Assessment of English as a foreign language vocabulary with a diagnostic complex test] *Iskolakultúra, 25*(12), 28-56.

Thékes, I. (2016). An empirical study into Hungarian Young learners' English as a foreign language learning strategies. *International Journal of Research Studies in Language Learning*, *8*(1), 1-22.

Vidákovich T., Vígh T., Sominé Hrebik O., & Thékes I. (2013). Az angol & német nyelvi szókincs 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.

Vígh, T., & Thékes, I. (2014). Results of a diagnostic English as a foreign language vocabulary test battery among Young learners. In E. Korom, & A. Pásztor (Eds.), *12th Conference on Educational Assessment* (p. 125). Szeged: SZTE BTK Neveléstudományi Doktori Iskola.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

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.

Publications related to the dissertation

Thékes, I. (2016). An empirical study into Hungarian Young learners' English as a foreign language learning strategies. *International Journal of Research Studies in Language Learning*, *8*(1), 1-22.

Thékes, I. (2015). An integrated diagnostic vocabulary test for young EFL learners. *Journal of Language Teaching and Linguistics, 6*(2), 189-209.

Thékes, I. (2015). Angol, mint idegennyelvi szótudás mérése egy diagnosztikus komplex online teszttel fiatal nyelvtanulók körében. [Diagnostic assessment of English and German as a foreign language vocabulary amongst 6th graders]. *Iskolakultúra, 25*(12), 28-56.

Vígh, T., S Hrebik, O., Thékes, I. & Vidákovich, T. (2015). Fiatal nyelvtanulók német és angol alapszókincsének online diagnosztikus vizsgálata. In B. Csapó & A. Zsolnai (Eds.), *Online diagnosztikus mérések az iskola kezdő szakaszában.* (pp. 13-35). Budapest: Oktatáskutató és Fejlesztő Intézet.

Vígh, T., & Thékes, I. (2014). Results of a diagnostic English as a foreign language vocabulary test battery among Young learners. In E. Korom, & A. Pásztor (Eds.), *12th Conference on Educational Assessment* (p. 125). Szeged: SZTE BTK Neveléstudományi Doktori Iskola.

Thékes, I. (2014). Assessment of Young learners' English as a foreign language vocabulary knowledge: Results of a diagnostic test with the Rasch model. In J. Horvath & P. Medgyes (Eds.), *Studies in honour of Marianne Nikolov* (pp. 87-100). Pécs: Lingua Franca Csoport.

Thékes, I. (2014). Egy kontrollcsoportos angol szótanulási kísérlet eredményei magyar főiskolások körében. [Results of an EFL vocabulary learning control-group experiment amongst college students] In J. Bárdos, L. Kis-Tóth, & R. Racskó (Eds.), *Új kutatások a neveléstudományokban. Változó életformák,*

régi és új tanulási környezetek [New research in education science. Changing lifestyles, old and new learning environments] (pp. 269-181). Budapest: MTA Pedagógiai Tudományos Bizottság.

Thékes, I. (2014). Egy integrált angol, mint idegen-nyelvi szókincsteszt eredményei 6. évfolyamosok körében. [Results of an integrated English as a foreign language test among 6th graders]. In E. Korom, & A. Pásztor (Eds.), *12th Conference on Educational Assessment* (p. 175). Szeged: SZTE BTK Neveléstudományi Doktori Iskola.

Thékes I. & Bányai, M. (2014). Egy fiatal nyelvtanulókat vizsgáló szótanulási stratégia-kérdőív fejlesztése. [Development of a vocabulary learning strategies questionnaire investigating Young learners]. In E. Korom, & A. Pásztor (Eds.), *12th Conference on Educational Assessment* (p. 176). Szeged: SZTE BTK Neveléstudományi Doktori Iskola.

Thekes, I. (2014). Assessment of Young learners' English as a foreign language vocabulary knowledge: Results of a diagnostic test with the Rasch model. In J. Horvath & P. Medgyes (Eds.), *Studies in honour of Marianne Nikolov* (pp. 87-100). Pécs: Lingua Franca Csoport.

Vidákovich T., Vígh, T. S. Hrebik, O., & Thékes, I. (2013). Az angol és német nyelvi szókincs 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.