

On the Evaluation of Lexical Profiles in the Iranian High School and University EFL Text books

Ali Asghar Eftekhary, Ph.D.

Islamic Azad University, Tehran South-Branch
ali.eftekhary90@gmail.com

Soudabeh Taheri, M.A.

I.A.U. Tehran-Iran

Received 2010 August 23 Accepted 2010 September 22

ABSTRACT

This research tried to investigate the vocabulary profiles in Iranian high schools and University textbooks. In order to do so, all the four Iranian high school textbooks which are in the form of a series called The Right Path to English were scanned onto a computer to form a data file composed of all the vocabulary items existing in the books. The data file was called corpus.

After that, the most frequent textbooks offered in Iranian universities as pre-university courses-given to students who ranked below 33% in the university entrance examination followed by General Service English courses which were two of the most frequently used books namely Reading for General English and Concepts and comments were selected and scanned onto a computer. The latter data file was compared against the corpus through RANGE software to seek the probable existence of a meaningful gap in the number of vocabulary items between the textbooks in high schools and universities in Iran. The obtained results proved the existence of such a gap which means the English words that students are exposed to during high school are not recycled in their English textbooks of university; therefore they face lots of problems in using these texts.

Key words: recycling of words, vocabulary gap , word frequency, vocabulary, text books, vocabulary profiles

On the Evaluation of Lexical Profiles in the Iranian High School and University EFL Text books

"Vocabulary is a core component of language proficiency and provides much of the basis for how well learners speak, listen, read, and write." Without an extensive vocabulary and strategies for acquiring new vocabulary, "learners often achieve less than their potential" and may be discouraged from making use of language learning opportunities such as listening to the radio, listening to native speakers, using the language in different contexts, reading, or watching television (Richards, 2002).

Among numerous problems lying in the way toward acquiring lexical items, one is exposure issue. Exposure basically relates to how often learners view, practice and use, a vocabulary item. The more often a learner encounters a vocabulary item, the higher the chances for him/her to acquire the word; therefore, the attempt is to expose the learners to the lexical items as often as possible. Language teaching in Iranian educational system begins in Junior High Schools with three books and then followed by four books in high schools. As Iranian students manage to go to universities they have to take English courses both in General Service and in the specific fields as well. Naturally/ students are exposed to foreign language in its different forms one of which is vocabulary items presented in various ways. But after these long years of exposure to foreign language, very few students achieve a high level of language proficiency.

Statement of the Problem

The place of vocabulary in L2 teaching and learning has deservedly received great attention in recent years. Due to this fact, materials designers have been trying to provide L2 learners with a good amount of exposure to new vocabulary items.

In Iran students are exposed to a number of vocabulary items in high schools. These words are usually presented in written form. When students start university they are exposed

to new English textbooks which means being exposed to more items. There are two issues to be considered. The first one is whether the lack of appropriate vocabulary knowledge is due to the number of vocabulary items, words, they know, or due to the insufficient exposure to these items.

Significance of the Study

Vocabulary learning is central to language acquisition, whether the language is first, second, or foreign. Despite the fact that vocabulary has not always been recognized as a priority in language teaching, interest in its role in second language (L2) learning has grown rapidly in recent years and specialists now emphasize the need for a systematic and principled approach to vocabulary by both the teacher and the learner.

This question, "Which words do learners need to know?" is where the significance of this study lies. Depending on goals of different learners and their levels, the type and number of lexical items to be taught, vary. It seems so hard a decision for syllabus designers as to which words are to be included and which ones are to be excluded. This is the fact that makes it important to know if the textbooks being used in Iranian education system both in high schools and universities expose students to lexical items generally of high frequency and range or not.

Research Questions

1. Since Iranian students are exposed to new vocabulary items both in high school and university as well, are the words that students learn in high school recycled in university GSE books?

2. Are Iranian students' problems in coping with word meanings due to the Gap between the books in high school and university? Or Is there a Gap between these two levels in regard with new words?

Research Hypothesis

Provided that the English words that students are exposed to during high school

studies are not recycled in their English textbooks; the following hypotheses can be stated:

1. The vocabulary profiles of the university textbooks are not statistically different from the high school textbooks in text-indexes such as type, type/token ratio, frequency of occurrence and, the like.

2. Iranian students' problem with vocabulary in academic levels does not lie in the insufficient recycling of words presented in the former levels.

Literature Review

Vocabulary is considered as one of the important components of language knowledge. Horwitz (1988) found that a large number of ESL students completing her questionnaire either agreed or strongly agreed that "the most important part of 'learning a foreign language is learning its vocabulary". Also research has shown that "lexical errors tend to impede comprehension more than grammatical errors, and native-speaking judges tend to consider lexical errors as more serious than grammatical ones" (Ellis, 1994). Widdowson, 1993(as cited in Lewis, 1993: 115) states that the more we consider the matter, the more reasonable it seems to suppose that it is the starting point, and that the syntax need to be put to the service of words and not the other way round. Interest in second language vocabulary acquisition (hereafter SLVA) has grown steadily in the last thirty years. It has given rise to the procedure of several excellent books (e.g., Huckin, Haynes, & Coady, 1993; Huckin & Coady, 1997; Nation, 1990; Schmitt & McCarthy, 1997; Singleton, 1999; Schmitt, 2000). A British study by Milton and Meara (1995) found that learners studying in foreign language environment could learn at a rate of 2500 words per year, while the rate for those studying in classrooms at home in the L1 environment was only about 550. A possible clue to the source of the variance in frequency estimates that different numbers of encounters are needed by different levels of learners.

Method

The project is a quantitative description of textbooks being taught in Iranian education system both in high schools and universities. The study was performed with the following specifications:

Corpus

To run this study, which is a corporal analysis of vocabulary items in Iranian high schools and university text books, the most necessary material is the corpus which is firstly composed of four versions_1 to 4_ of the series *The Right Path to English* (Birjandi, et al.). The series is Iranian Educational system's textbook for students of high school in four successive years.

Iranian university textbooks (in Azad University) offered as pre-university course, which is presented to students who scored below 33% in Azad University entrance exam. This indicates that these participants are not competent enough to move forward to continue their English studies. In this section, two of the most frequently used books namely *Reading for General English* (Pourgive et al., 2009) and *Concepts and comments* (Ackert, 2005) are used.

Instrumentation

this study was done in two phases. The first phase (hereafter called the corpus phase) was to develop the necessary body of text –corpus. In order to develop such a body first a scanner was used to scan the books on a PC. Then by a software known as OCR (Optical Character Recognition), the format of the scanned books was converted to text which would be a necessary format for Word Smith family of software. In the second phase the software *RANGE* was used to analyze the body developed in the corpus phase. *RANGE* is used to compare the vocabulary of up to 32 different texts at the same time. For each word in the texts, it provides a range or distribution figure (in how many texts the word occurs), a headword frequency figure (the total number of times the actual headword type appears in all the texts), a family frequency figure (the total number of times the word and its family

members occur in all the texts), and a frequency figure for each of the texts the word occurs in. *RANGE* can be used to find the coverage of a text by certain word lists, create word lists based on frequency and range, and to discover shared and unique vocabulary in several pieces of writing.

The word lists available for RANGE

Three ready made base lists are available. The first (BASEWRD1.txt) includes the most frequent 1000 words of English. The second (BASEWRD2.txt) includes the 2nd 1000 most frequent words, and the third (BASEWRD3.txt) includes words not in the first 2000 words of English but which are frequent in upper secondary school and university texts from a wide range of subjects. All of these base lists include the base forms of words and derived forms. The first 1000 words thus consists of around 4000 forms or types. The sources of these lists are *A General Service List of English Words* by Michael West (Longman, London 2002) for the first 2000 words, and The Academic Word List by Coxhead (2000) containing 570 word families. The first thousand words of A General Service List of English Words are usually those in the list with a frequency higher than 332 occurrences per 5 million words, plus months, days of the week, numbers, titles (Mr, Mrs, Miss, Ms, Mister), and frequent greetings (Hello, Hi etc).

Procedure

Having assured that the above mentioned university textbooks were the most frequently used ones, they were proven to be the most repeatedly offered textbooks. All these books had to be converted to computer files by means of an optical scanner and one of OCR (optical character recognition) software to fit *RANGE* necessary format. This composed the corpus against which text books were to be compared. After that, each single book was analyzed by *RANGE* to mark the most frequent words in each book and they were compared against the software base lists to see if the words exist in the list.

Next step was to compare university books

against the corpus composed of high school textbooks which was about to check if the hypothesis put forward were correct or not.

Results

Before presenting the data in tables, it seems necessary to give a short definition of the head words in the tables. The first one is Word List, which is shown in table 1-8 and consists of the three lists of most frequent 1000, 2000 and 3000 words by Nation et.al available for RANGE. Tables 7 and 8 show the Word List of the corpus which consists of the words used in Iranian high school textbooks. In linguistics, a distinction is sometimes made between classes of linguistic items and actual occurrences in speech or writing. The class of linguistic units is called *type* and the examples or individual members of the class are called *token*.

Table 1 shows that 13968 of the running words in book one in high school are in base list one and these 13968 words make up 84.48% of the total running words in the text (16534). These words come in 957 types which covers 58.68% of the words in base list one and they all belong to 549 word families. In the word list, column one, two, three refer to each of the base lists. 1490 of the words of book one are in list two which is 9.01% of the number of words in the running list, 290 types that is 17.78% of the types in the list and in 207 word families. 105 words of the book cover up to 0.64% of the tokens in list three, 34 types that is 2.08% of the third list and 29 families. 5.87% of the words (i.e. 971 tokens) from 350 types are not in any of the three base lists.

Word list	Tokens	%	Types	%	Families
One	13968	84.48	957	58.68	549
Two	1490	9.01	290	17.78	207
Three	105	0.64	34	2.08	29
Not in the Lists	971	5.87	350	21.46	—
Total	16534	—	1631	—	785

Table 1 book one compared against base lists

Table 2 indicates the information presented above about the book two in high school compared against the base lists. 7994 of the words in this book appear in base list one. They belong to 799 types in 508 word families that cover up to 82.00% of the total words which overlaps with 49.69% of the types of the base list. 817 (i.e. 8.38% of the words in book two) is the number of words in book two that exist in

base list two. They are from 243 types of lexical items which is 15.11% of the sum of words in book two appearing in 189 word families. What book two has in common with base list three is 914 tokens 9.38% belonging to 7 types which is 0.44% of book two from 6 families. 914 words from the book are not in any of the base lists that are 9.38% of the words and 559 types being as 34.76% of types.

Word list	Tokens	%	Types	%	Families
One	7994	82.00	799	49.69	508
Two	817	8.38	243	15.11	189
Three	24	0.25	7	0.44	6
Not in the Lists	914	9.38	559	34.76	—
Total	9749	—	1608	—	703

Table 2 Book two compared against base lists

Table 3 reports the results of analyzing book three. 11455 of tokens in book three overlap with the tokens in base list one, 1207 of tokens with base list two and 313 ones with base list three which respectively make up 82.46%, 8.69% and 2.25% of the total number of vocabulary items in book three. 1076 (58.57%)

types of items in book three (in 620 word families) exist in base list one, 307 (16.71%, in 223 word families) in base list two and 87 (4.74%, in 62 word families) in base list three. 916 of the words in book three do not exist in any of the base lists which are in 1837 types and 905 word families.

Word list	Tokens	%	Types	%	Families
One	11455	82.46	1076	58.57	620
Two	1207	8.69	307	16.71	223
Three	313	2.25	87	4.74	62
Not in the Lists	916	6.59	367	19.98	—
Total	13891	—	1837	—	905

Table 3 book three compared against base lists

The last book in Iranian high schools is called pre-university course text book (hereafter book four). Table 4 shows the RANGE outcome for book four. From the total 17861 words in this book, from 3040 types and 1088 families, 13761 tokens meaning 77.04% appear in base list one. 1233 tokens that is 6.90% exist in base list two. 803 tokens being as 4.50% of the running words in book four are in base list three.

The tokens existing in list one are from 1350 types from 705 families, 384 ones in base list two from 258 families and 200 in base list three from 125 word families which respectively cover 44.41%, 12.63% and 6.58% of the total number of words in book four. 2064 words from 1106 types of words in book four do not exist in any of base lists which is 11.56% of the total number of words.

Word list	Tokens	%	Types	%	Families
One	13761	77.04	1350	44.41	705
Two	1233	6.90	384	12.63	258
Three	803	4.50	200	6.58	125
Not in the Lists	2064	11.56	1106	36.38	—
Total	17861	—	3040	—	1088

Table 4 Book four compared against the base lists

A short look at the tables reveals the fact that as Iranian students in the first year are exposed to 16534 vocabulary items and then surprisingly in the second year, they come across 9749 words. Then in the third year, they are exposed to 13890 lexical items. When book four takes its turn, 17861 words appear. Apparently there is no logical order in choosing which/how many words to be taught. Tables 5 and 6 report the output of RANGE for

the books in university. The first thing that catches one's eyes in these two tables is the total number of words in these books which is 44361 words in the first one, *Concepts and Comments*, and 25858 words in the second one, *Reading for General English*. This difference in the number of lexical items might now be of no significance. A later comparison of these two books against the corpus in this study revealed that if this gap is meaningful or not.

Word list	Tokens	%	Types	%	Families
One	35187	79.32	1691	43.56	791
Two	3726	8.40	679	17.49	394
Three	909	2.05	177	4.56	111
Not in the Lists	4539	10.23	1335	34.39	—
Total	44361	—	3882	—	1296

Table 5 *Concepts and Comments* compared against the base lists

Word list	Tokens	%	Types	%	Families
One	20462	79.13	1660	45.45	840
Two	1907	7.37	675	18.48	449
Three	840	3.25	303	8.30	199
Not in the Lists	2649	10.24	1014	27.77	—
Total	25858	—	3652	—	1485

Table 6 *Reading for General English* compared against the base lists

Following this phase, the previously mentioned corpus- the body of data consisting of all the words in high school text books- was set as the base list in RANGE and the university text books were compared against this body to answer the research questions in this project. The results are reported in tables 7 and 8. Table 7 shows the results of juxtaposing the first book offered in universities, *Concepts and Comments*. From the total number of 44361

tokens in this book, 38806 appear in the corpus which means 87.48% of the tokens in high school text books. They are derived from 2062 types (53.12% of the types in the base list) and are from 2062 word families. 5555 words that is 12.52% of the words in the first text book in universities do not appear in the corpus. These words are from 1820 types. The total number of types and word families were calculated as 3882 and 2062.

Word list	Tokens	%	Types	%	Families
The corpus	38806	87.48	2062	53.12	2062
Not in the Lists	5555	12.52	1820	46.88	—
Total	44361	—	3882	—	2062

Table 7 *Concepts and Comments* compared against the corpus

After the first book in Iranian universities, offered as pre-university course, the second book which is compulsory for students of all majors takes its place. The most frequent offered book is *Reading for General English* (Pourgive et al, 2009). Table 8 indicates that the total number of words in this book is 25858 which belong to 3652 types and are from 1825

word families. 22016 of these words which mean 85.14% of the total words appear in the corpus. This translates 49.97% of the total types (i.e. 1825 types) in the corpus. They are from 1825 families of lexical items. 3842 words do not exist in the corpus. That means 14.86% of the total. This is 1827 types and 50.03% of the overall types of the words in the book.

Word list	Tokens	%	Types	%	Families
The corpus	22016	85.14	1825	49.97	7825
Not in the Lists	3842	14.86	1827	50.03	—
Total	25858	—	3652	—	1825

Table 8 *Reading for General English* compared against the corpus

Discussion

Chart 1 shows the total number of running words in the corpus, book one and two in university. As can be seen in the chart the number of words in university in the books decreases as students

make it to the second book. In book one they are exposed to an overall number of 44361 words while in book two this figure falls to 25858 which is a considerable number.

books 1 & 2 against the corpus

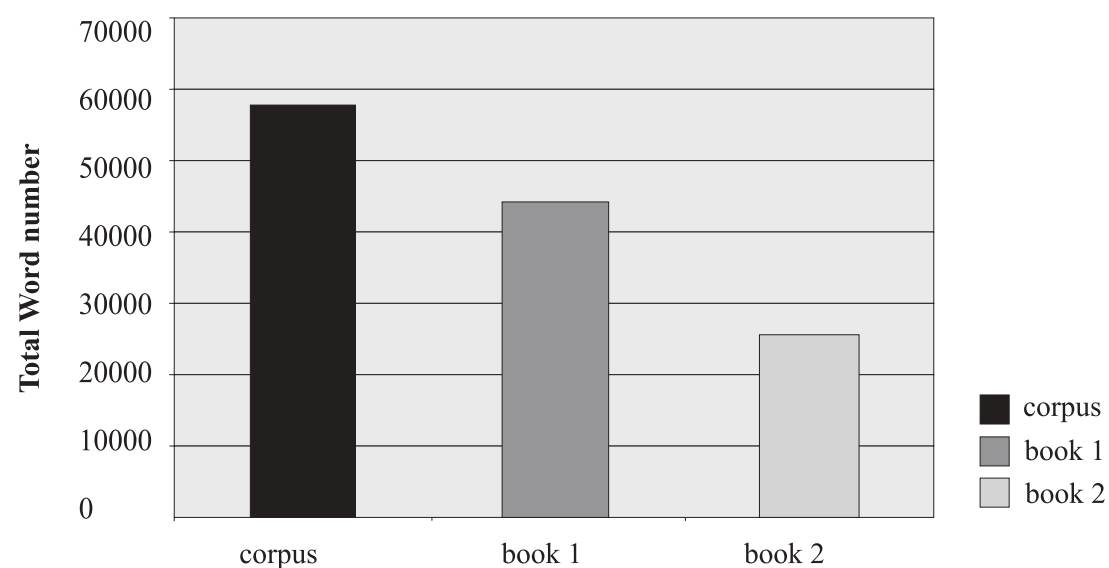


Chart 1 The Total Number of Words in University Textbooks

Based on the data obtained from the statistical analysis of the university text books against the ones in high schools, the hypotheses of this study turned out to be true. The tabulated data depicted that the previously mentioned gap, really exists and seems to be a meaningful one which hinders vocabulary learning for students in Iranian educational system.

Conclusions

Considering language and language learning, every one has experienced that the more often a learner encounters a vocabulary item the higher the chances for him/her to acquire the word. This concept is technically called Exposure issue.

Exposure issue basically relates to how often learners view, practice, and use a vocabulary item. This study tried to find if the new vocabulary items taught in four years of high schools in Iran are recycled in higher levels of education or not. In other words the attempt was made to see if an Iranian high-school graduate has the chance to be exposed to the words already learned at high school or he/she is exposed to completely new lexical items without previous words being recycled which in turn produces a gap between the words taught in high school and university text books. This study tried to question the existence of this gap.

As it was shown in tables 1-6 there is not much of a difference in the books concerning the percentages reported for each book's coverage of the base lists. However, a brief look at the total number of words in each book reveals that there doesn't seem to be a logical rationale for deciding on the number of words to appear in each book. As shown in the tables, the number of words in book one in high school is 16534. Book two contains 9749 words. Keeping in mind that all the first three books have the same number of lessons, this fall which is by a considerable figure ($16534-9749=6785$) is a bit strange. Then in book three, this figure rises to 13891 and the biggest figure of all belongs to book four, 17861. Moving on to text books in university, the first one, known as *Concepts and Comments* which is offered as pre university

course to those students who scored below 33% of the questions in university entrance exam and is supposed to be a gap-filler between high school and university general service courses, contains 44361 vocabulary items while the second book offered as general service English has 25858 words. Of course at this level this gap cannot be interpreted as meaningful, but as this study moved forward to comparing these two text books in university with the corpus, composed of all the four books in high school, this gap can be interpreted as meaningful.

The pre-university course in Iranian universities (especially Islamic Azad University) is originally designed to help students with a weak performance on the entrance examination to escalate their level of language proficiency, in other words fill the gap in their language proficiency level, to fit the general service course level. As it turned out, concerning the number of lexical items taught, this study results showed that: first as previously mentioned the total number of words in *Concepts and Comments* was way more than the running words in the book *Reading for General English* (i.e. 18503). It seems that the so called gap-filler not only does not cover the gap but also even expands it. Second, as it was shown in table 7 the number of words that do not appear in any of the text books in high school is 5555 which is 12.52% of the total words in the book. What makes it worse is that these words are mostly from the most frequent words in the book.

Having compared *Reading for General English* against the corpus, results showed that 3842 words that are 14.86% of the book did not overlap with the corpus. Like the words in book one, these words are unfortunately among the most frequent lexical items in the book. Having the percentage figures mentioned above and the figures reported in tables 7 and 8 for the token coverage percentage, one might ask: considering these figures, isn't a student much likely to comprehend the text or isn't a student able to guess the meaning of the words from the context? To answer this very pertinent question the researchers would like to say: Since learners depend on vocabulary as their

first resource; thus a rich vocabulary makes the skills of listening, speaking, reading, and writing easier to perform.

Therefore, there has been continuing interest in whether there is a language knowledge threshold which marks the boundary between having and not having sufficient language knowledge for successful language use (Nation 2001: 144). Historically, experienced teachers such as West (1953) considers one unknown word in every fifty words to be the minimum threshold necessary for the adequate comprehension of a text. Others such as Johns (as cited in Bensoussan and Laufer 1984) considered 95% 'coverage,' or one unknown word in every twenty words, to be the threshold, which was later confirmed by Laufer (1989). Hu and Nation (2000) conclude that for largely unassisted reading for pleasure, learners would need to know around 98% of the running words in the text. However, the current thinking in the field of vocabulary teaching and learning puts the threshold of meaningful input at 95% (Schmitt and McCarthy 1997, Tono *et al.* 1997, Read 2000 and Nation 2001. Nation (2001) assures us that "if more than five percent of the running words are unknown, then it is likely that there is no longer meaning-focused learning because so much attention has to be given to language features" (pp. 388-389). Clearly, it is first necessary to examine what vocabulary exists in learners' textbooks, and to determine if the learners are able to meet the 95% comprehension criteria. If not, educators must then provide the supplemental vocabulary to bridge this gap. Without this kind of bridge, learners would face a boring amount of dictionary work. A glance at the tables 7 and 8 shows that the number of known words (the percentage figure reported for the number of existing tokens in the corpus) in *Concepts and Comments* covers 87.48% of the corpus in this study which means out of each twenty words 3 are unfamiliar. Also, table 8 shows that 85.14% of the running words in *Reading for General English* are known for students. This means out of each twenty words 4 are unfamiliar. These figures are less than 95% and this is the fact that makes it difficult, if

not impossible, for students to guess the meaning of the words from the context.

As it comes to conclusion one could say that because the number of frequent words in university books is way more than the number of running words in high school, apparently the gap introduced in the research hypothesis exists. Also, it seems that the gap-filler, pre-university course at university, fails to meet the purpose it was designed for and seems to even add more to the difficulty for students.

References

- Ackert, P.(2005). Concepts and Comments: a reader for English as a second language
 Birjandi, P,etal. (2009). The Right Path to English, Book One, Two, Three and Four. Ministry of Education, Tehran.
 Coxhead, A. (2000). A New Academic Word List. TESOL Quarterly, 34.2: 213-238. Department of English, The University of Tokyo (1998) The Hartly, B., Viney, P. (1979-85). Streamline. Oxford: Oxford University Press.
 Ellis, R. (1994). The study of second language acquisition. Oxford: Oxford University Press.
 Horowitz, L.M. (1988) Associative symmetry and second language learning. Journal of Educational Psychology 63, 3:287-294.
 Hu, M. and P. Nation (2000). Unknown vocabulary Density and Reading Comprehension, Reading in a foreign Language, 13.
 Huckin, T. & Coady, J. (1997). Second Language Vocabulary Acquisition (pp. 203-224). Cambridge: Cambridge University Press.
 Huckin, T., Haynes, M., & Coady, J. (Eds.). (1993). Second language reading and vocabulary learning. Norwood, NJ: Ablex.
 Laufer, B. (1998). The lexical plight in second language. In J. Coady, & T. Lewis, M. (1993), The lexical

approach. Hove, England: LTP.

- Nation, P. (1990). Teaching and learning vocabulary. New York: Newbury House.
 Nation, P. (2001). Learning Vocabulary in another Language.
 Pourgive, F., Sadeghi, F. and Yamini, M. (2009). Reading for general English. Tehran: SAMT.
 Read, J. (2000), Assessing vocabulary. Cambridge: Cambridge University Press.
 Richards, J. (2002). Methodology in Language Teaching. New York: Cambridge University Press.
 Schmitt, N., & McCarthy, M. (Eds.). (1997). Vocabulary: Description, acquisition, and pedagogy. Cambridge: Cambridge University Press.
 Schmitt, N. (2000). Vocabulary in language teaching. Cambridge: Cambridge University Press.
 Singleton, D. (1999). Exploring the Second Language Mental Lexicon. Cambridge University Press, Cambridge.
 Tono, Y. (1997) Eigo Goi Shuutokuron (1997) (Theories of Teaching and Learning English Vocabulary). Tokyo: Kagensha. (In Japanese)
 West, M. (1953). A general service list of English words. London: Longman Green.



Ali A. Eftekhary

is an assistant professor holding a Ph.D from the university of Texas in Austin, Texas in English and Persian as a foreign language; minor Middle Eastern Studies. Teaching Persian for fourteen years in the University of Texas at Austin. He has been teaching in Islamic Azad University since 1991. He is the author of ABC in Education Research.



Soudabeh Taheri

has a B.A in English Translation from Khorasan Islamic Azad University and an M.A in ELT from Islamic Azad University, South Tehran Branch. She has started teaching in different institutes, since 2002 in Isfahan and has worked as a translator in Spehr Translation Office since 2004.