The Vocabulary Profile of Iranian English Teaching School Books

Hadi Farjami [*]

Assistant Professor, Semnan University, Semnan, Iran

Abstract

This paper provides a fairly detailed corpus-based vocabulary profile of the Iranian EFL books used in public schools. To this end, the WordPerfect files of all the seven books were converted to text format to get rid of the formatting features and be compatible with the software used for analysis. The software tools used were the Compleat Lexical Tutor suite, version 6.2 (Cobb, 2011), AntConc (Anthony, 2012), and AntWord Profiler (Anthony, 2012). The output of the analysis included general counts of words in Iranian school books at different levels, the frequent function and content words, frequent n-grams, frequent metalinguistic words, the coverage of several well-known, corpus-based word lists in these books, the range of the words across the books, and the amount of vocabulary recycling. The paper discusses the vocabulary representativeness and recycling and the adequacy of exposure to English in these EFL books. Detailed word frequency tables as well as some practical implications of the quantitative results constitute important features of this article.

Keywords: Iranian ELT schoolbooks; Vocabulary input; Vocabulary profile; Vocabulary recycling; Word frequency; Word lists

---

[*] *Corresponding address:* Department of English Language and Literature, Semnan University, Semnan, Iran. Postal Code: 3513119111
E-mail address: hfarjami@semnan.ac.ir

Introduction

Having a clear picture about the quantity of vocabulary input in textbooks is particularly important in a foreign language learning setting, where there is a serious shortage of natural exposure to the target language and, hence, its learning, to a large part, depends on textbooks. In partial response to this need, this study provides a quantitative profile of the vocabulary in the seven EFL textbooks which were prepared and officially assigned to schools to teach by the Iranian Ministry of Education. Considering the uncontroversial role of vocabulary as an essential component of input in language learning (e.g., Lessard-Clouston, 2013) and the central role textbooks play as the medium of that vital input (e.g., Richards, 2005), the study seems significant and of much practical value. Its results furnish awareness in quantitative terms about how much vocabulary Iranian learners are exposed to and how frequently that vocabulary is practiced. This, in turn, helps us see what is missing in our teaching practice and what should be emphasized more.

The study reported in this article used corpus linguistics methodology, as a usual modern mode of lexical exploration, to get detailed counts at different levels of the words in Iranian English teaching schoolbooks, and obtain information about the frequency of function and content words, the frequent collocations, the range of words across the books, and the rate of introducing new words and their recycling. Moreover, comparisons were made between the vocabulary in these books and some previously researched word lists, i.e., the General Service List, the Academic Word List, Longman Communication 3000 Words, and Ogden's Basic English Words. When language teaching practitioners are aware of the degree to which the vocabulary in these textbooks correspond with these research-based lists, they will be in a position to plan their teaching with information about the vocabulary which is neglected or downplayed in these books and should be remedied.

Review of the Related Literature

Studies on Vocabulary Input Quantity

Vocabulary size has been found to be a predictor of reading comprehension (e.g., Keshavarz & Mohammadi, 2009; Nation, 2006). Concerning vocabulary size, Laufer (1997) suggests 3000 word families as the minimum for reading in English. Zahar, Cobb, and Spada (2001) consider this functional vocabulary

knowledge necessary for reading comprehension, but they emphasize a vocabulary of 5,000 word families for effective reading and understanding. Later research by Laufer and Ravenhorst-Kalovski (2010) set a higher standard. They suggested two thresholds of vocabulary knowledge for reading comprehension: optimal knowledge, i.e., knowing 8,000 word families and covering 98% of the text to read and minimal knowledge, i.e., knowing 4,000–5,000 word families and covering 95% of the text. Research by Prichard and Matsumoto (2011) confirmed this minimum requirement as their data confirmed that many participants in the 90-95% coverage range had difficulty comprehending texts.

Another question is the number of repetitions and amount of recycling, that is, the number of times a word must be encountered. There is less agreement over this question because depending on factors such as learning criteria, conditions of learning, learners' levels, individual differences as well as quality and context of exposure, researchers have come up with different figures concerning optimal repetition of vocabulary. Estimates range from six (Saragi, Nation, & Meister, 1978) to 20 (Herman, Anderson, Pearson, & Nagy, 1987). Saragi, et al. (1978) found that words presented to learners fewer than six times were learned by half of the participants, while words presented six times or more were learned by 93%. Nation (2001) suggested 16 encounters was a common figure in the literature. Although they make no definite statement about the amount of exposure needed to learn a word, Brown, Waring, and Donkaewbua (2008) emphasize that repetition does have an effect and the greater the number of occurrences of a word, the higher the likelihood of its learning and recall (See also Rott, 2007).

Research has shown that the rate of vocabulary learning is not static (e.g., Laufer & Rozovski-Roitblat, 2011; Reznick & Goldfield, 1992). For example, more advanced learners are less dependent on frequency and, therefore, learners' level should be taken into consideration in studying optimal recycling of vocabulary. In the study by Zahar et al. (2001), mentioned above, it was shown that frequency was more than three times as determined for learners scoring 50% on their level test than for learners scoring 60%. The researchers argued that the number of occurrences needed for acquisition drops considerably over the course of learning the second thousand words. Moreover,

in an elaborate study of the effect of word recycling by Laufer and Rozovski-Roitblat (2011), occurrence was found to have an effect on retention, but the effect of task-type was stronger.

A related and complicating issue to the quantity and frequency of exposure is the mode and strategies for vocabulary learning. For example, while many researchers acknowledge the role of explicit instruction and practice in vocabulary learning (e.g., Schmitt, 2008; Walters, 2006), many others give more emphasis to incidental acquisition of vocabulary (e.g., Horst, 2005; Huckin & Coady, 1999; Krashen, 1993a). A staunch advocate of the latter position is Krashen, who in numerous publications supports frequent exposure through what he calls free voluntary reading, extensive reading or sustained silent reading (e.g., Krashen, 1989, 1993a, 1993 b). In a recent article (Krashen, 2012), he reviews empirical research on the role of frequent exposure in vocabulary development and upholds the position that "direct instruction cannot deal with the size and complexity of vocabulary learning" (p. 33) because there are too many words to be acquired and it is impossible to learn the subtle and fuzzy meanings of words through planned instruction.

Corpus Studies and Word Lists
It makes pedagogical sense to use frequency as a guide in vocabulary teaching and to know which words learners should focus on during their limited time (Nation, 1990). In fact, Sinclair (1991) noted that "anyone studying a text is likely to need to know how often each different word form occurs in it" (p.30). Moreover, it is through frequent exposure that form-function and form-meaning associations are formed and consolidated. The Power Law of Practice is the fact that there is a relationship between frequency of trials and learning (DeKeyser, 2001). This classic fact about language learning has made syllabus design and classroom methodology for language teaching major fields of data application provided by corpus-linguistic research (Ganger & Brent, 2004). Corpus research, which attempts to discover patterns associated with lexical and grammatical features (Flowerdew, 2001), has influenced syllabus design and methodology in English teaching in two major ways: 1) providing descriptions of the target language and thus affecting the content of what the teacher teaches and, 2) producing attested language teaching materials (Hunston, 2002). Among other contributions, corpus linguistic research has provided information

concerning the frequency and ranking of words in general and specialized English corpora.

According to Nation (1990), the 4,000-5,000 most frequent words account for 95 percent of written English and 85 percent of speech consists of the 1000 most frequent words. So, it makes pedagogical sense to use frequency as a guide in vocabulary teaching and know which words learners should focus on during their limited time.

Edward Thorndike was the first to document word frequencies and compile a teaching-oriented word list. He manually counted the frequency of 18,000,000 English words in educational texts and published a series of word books for teachers culminating in *A teacher's word book of 30,000 words* (Throndike & Lorge, 1944). Michael West (1953) compiled the General Service List (GSL), which contains the most widely used 2000 English words and has had a wide influence by serving as the basis for graded readers and other texts. The GSL has remained valid in terms of coverage and frequencies so much that its first 1,000 words covers 72% of the texts in the Brown corpus, a diverse corpus of over 1,000,000 words. More recent lists of high frequency English words include the Oxford 3000™, Longman Communication 3000, and the Academic Word List (AWL). Coxhead (2000) compiled the well-known AWL from a corpus of 3.5 million written academic words outside the first 2000 most frequent English words.

Although the relationship between frequency on the one hand, and language use and acquisition on the other hand cannot be denied or ignored, frequency is not everything in learning. For example, there is a generally reverse connection between the frequency of a linguistic item and its complexity: Complex structures are less frequent than simpler ones (Roeper, 2007). This means that frequency alone does not determine the learning needs of learners because it is possible to be familiar with most of the words in a text and still have very little understanding of the content (Milton, 2009). Thus, the most frequent does not necessarily mean the most useful; in some cases, other words may be more urgent to learn for comprehension. As another complication for incorporating corpus linguistic techniques in teaching, some scholars have mentioned that teacher may be reluctant to employ corpus strategies in vocabulary teaching

(e.g., Romer, 2010) or may fail to develop effective corpus-based activities (e.g., Heather & Helt, 2012).

Vocabulary and Textbooks
In some situations, textbooks are "the basis for much of the language input learners receive and the language practice that occurs in the classroom" (Richards, 2005, p. 239). So, it has been relevant to examine the vocabulary of textbooks to learn about their content validity. For example, Sutarsyah, Nation, and Kennedy (1994) investigated the academic vocabulary load of books in different academic fields and found that the number of shared vocabulary items among specific fields was limited, that is, many of the words in each discipline were unique to that field. They indicated that "EAP courses that go beyond the high frequency academic vocabulary are of little value for learners with specific purposes" (p. 34). Scholfield (1991) looked at the rates of vocabulary introduction and recycling in EFL coursebooks and found significant differences. Nation and Wang (1999) investigated graded readers and concluded that most schemes are not well-designed in terms of vocabulary size. Milton's (2009) comprehensive study of vocabulary acquisition discusses, among other issues, the role of textbooks in teaching and learning vocabulary.

A recent review of studies on Iranian ELT schoolbooks by Riazi and Mosalanejad (2010) indicates that few studies have considered their vocabulary. In their comprehensive review, there is no mention of the studies done on textbooks at vocabulary level. The search by the present researcher did not reveal pithy studies, either.

The Study
The purpose of this study was to furnish a detailed quantitative picture of lexical input in the main stream Iranian ELT textbooks. The study targeted the following questions about Iranian English Teaching schoolbooks (IETSs):
  1. What are the more frequent function words in IETSs?
  2. What are the more frequent content words in IETSs?
  3. What is the rate of introducing new vocabulary items across IETSs?
  4. What are the more frequent n-grams in IETSs?
  5. What meta-linguistic words are frequently used in IETSs?

6. How representative of English lexis are IETSs based on comparisons with established word lists?
7. What is the range profile of the words across IETSs?

Iranian ELT School Books
Seven textbooks constitute a central element in Iranian English teaching in schools, which officially starts at the beginning of guidance (middle) school and stretches to pre-university studies. These books can be downloaded at *www.chap.sch.ir*. The books are illustrated with hand drawings except for the pre-university book which is mostly illustrated with snapshots. They all include inventories of vocabulary and grammar items for each lesson as well as general lists of words, phonetic symbols, and irregular verbs at the end of the books. Each book, except for the first one in the series starts with a few exercises to review the previous book. Although many learners come to the middle school with some familiarity with English, the first book in the series does not assume any familiarity with English and includes several lessons of alphabet practice. Lessons in the middle school books start with topical conversations which serve as springboards for sentence drills and grammar practice, sentence completion, and sentence production in writing and orally. Book 3 features a reading section toward the end of each lesson.

The lessons in the high school books are organized around passages, preceded by preparatory vocabulary exemplifications and followed by three types of comprehension questions. The lessons continue with oral or written sentence transformation exercises. There is a brief task targeting a communicative function in each lesson. There is also some minimal phonetic and vocabulary practice toward the end of each lesson. Although grouped as "speaking" and "writing" activities, discrete grammar practice is an integral part of high school books.

The eight lessons in the pre-university book constitute a course of reading and grammar with pre-reading questions and passages followed by comprehension questions, vocabulary, morphology, grammar exercises, and reading skills practice. Each lesson also includes a few questions for class discussion.

The Corpus and Data Analysis Software
The research reported here aspired to describe the lexical dimension of the seven EFL textbooks prepared by the Iranian Ministry of Education for middle (called "guidance" in Iran) and high schools. The description is based on corpora made up of the verbal information in the latest versions of these books, encoded as separate WordPerfect files. Having ensured that the files were free from spelling errors and irrelevant data, particularly in the case of copy-pasted material, the Word files were converted to **text** format to make them compatible with the intended data processing software.

Marginal information such as tables of contents, headers and footers, page numbers, and lists at the end of the books were not included in the analysis. To give a comprehensive and detailed picture of vocabulary in these books, some analyses were done in two steps, first including the metatextual information—instructions, headings, end of the lesson word lists, grammar summaries—then, applying the same analyses using a reduced corpus excluding metatextual material. Sixteen text banks were prepared for analysis:
1. Guidance School 1 (G1) (with and without metatext)
2. Guidance school 2 (G2) (with and without metatext)
3. Guidance school 3 (G3) (with and without metatext)
4. High School 1 (H1) (with and without metatext)
5. High School 2 (H2) (with and without metatext)
6. High school 3 (H3) (with and without metatext)
7. Pre-university (Pre-Uni) (with and without metatext)
8. The amassed full text of the seven books (with and without metatext)

The word lists used in the analyses included:
1. The GSL
2. The AWL
3. Longman Communication 3000
4. Ogden 850 Basic English Words
5. Function Words

Three software packages were used: Compleat Lexical Tutor (Cobb, 2011) for its text comparing tool, AntConc 3.3.5w (Anthony, 2012) for its frequency, word-list and N-gram tools, and AntWord Profiler 1.3.1 (Anthony, 2012) for its Range program. Although each of these applications provide a wide range of

tools, applications from different packages were used because different degrees of friendliness they showed with different analyses were seemingly the more reliable output for particular purposes, according to pilot analyses with small data sets.

Results

It is not practical to provide the flood of data which poured out of this analysis because of reasons having to do with space, organization, and clarity. So, only a selection of the data is tabulated below in response to the research questions above.

Table 1 provides the lexical variation and the number of tokens, types, and families both for function and content words. Lexical variation (LV) ratios show the diversity of words in texts. The higher a ratio, the fewer repetitions there are. The first textbook in the series includes 3,555 tokens, which boil down to 188 families of words, 49 of which are function words and 73 (188-115) of which are used only in the instructions or other metatextual places of the book. The last book in the series_ *English for Pre-University Students_* includes 1,277 families, realizing in 16,190 words. There are still 54 words which are only used metatextually. The majority of these words are content words as most function words are introduced in the earlier levels. In fact, the last two books have either no or few new function words.

The books together include 79,359 words comprising 1,877families. The increase in the number of words in the books does not seem to be balanced, whether the issue is considered based on a total count or without the metatextual words. For example, *H3* includes only 30 words more than *H2*, and there are 109 more words in H2 than in H1. Given the number of new words in these books, it is obvious that this rate of addition is not balanced. Such slight addition also underestimates the vocabulary learning capacity of learners after four years of language learning and cognitive growth.

The differences between the counts before and after the removal of metatextual words reveal the extent to which instructions and peripheral texts are used as opportunities for vocabulary teaching. There is a difference of 73 and 50 words for *H1* and *H2*, respectively, but the differences drastically shrink

as we move up to higher levels. This shows that instructions and peripheral texts do not provide systematic opportunities for vocabulary building.

Lexical variation of the books may also be worth considering. One way to measure this feature is to calculate the type-token ratios. High variation in instructional materials may mean insufficient practice and recycling. But, what level of variation is educationally sound should be empirically established. The lexical variation of IETSs steadily increases from level one to seven and almost doubles in the final stage. Low lexical richness at lower levels is partly due to the introduction and practice of function words.

Table 1
The Frequency Profile of Vocabulary in IETSs

Books	All tokens	All types	All families	Lexical variation ratios	Function words (total: 164 families)			Tokens without metatext	Types without metatext	Families without metatext	Lexical variation ratios
					Tokens	types	Families				
G1	3555	274	188	7.71	358	59	49	2742	182	115	6.63
G2	7677	534	382	6.95	462	112	76	5660	494	299	8.73
G3	9907	1077	591	10.87	532	133	98	7864	828	510	11.12
H1	15414	1492	886	9.67	598	171	127	12643	1406	806	11.12
H2	14171	1597	995	11.27	616	198	137	11517	1458	905	12.66
H3	12445	1656	1025	13.31	610	190	137	9982	1525	936	12.28
Pre-Uni	16190	2203	1277	13.60	624	195	142	13283	2104	1223	15.84
All Book s	79359	3790	1877	4.77	40846	235	153	63691	3535	1804	5.55

Table 2 presents function word families in the most frequent 200 IETS words. Quirk (1985) was referred to and Text-Compare tool in Compleat Lexical Tutor (Cobb, 2011) was used in extracting this information. In these 200 families, 76 (38%) are function words—49 from the first 100 families, 27 from the second. Table 2 also displays some function words beyond *would* (the 200th word) to give a taste of function words at lower frequencies. Obviously,

function words with frequencies lower than 50 are few. Numerals account for 7.21% and definite and indefinite articles account for 9.14% of the tokens.

Table 2
The Most Frequent Function Words in IETSs

the (1st) 4498	at 521	one 212	before 116	
be 3646	on 464	from 206	so 113	here 70
a 3220	can 417	all 186	any 112	nine 67
you 1807	for 411	as 184	five 112	than 67
to 1605	with 378	every 180	may 104	must 66
do 1550	there 366	if 180	lot 103	eight 63
he 1528	not 335	get 176	more 103	would (200th) 63
in 1466	no 321	where 170	because 99	seven 60
I 1382	will 309	but 158	which 99	ten 58
this 1205	or 297	should 148	too 98	into 54
it 1058	how 284	two 143	other 97	then 54
and 1031	some 276	up 141	much 94	never 52
they 1027	about 266	by 138	down 91	only 52
of 935	very 237	why 133	three 87	next 47
have 933	yes 237	could(100th) 128	four 85	each 46
what 728	when 234		six 73	few 46
she 691	who 234	out 124	first 71	over 46
we 598	many 216	after 119	most 71	

Of the 281 function types, indicated by Quirk (1985), 237 types were shared by the 7 books, which means that 44 types are not in the books including *ought, anywhere, beneath, hence, thence, towards, hither, nearby, nowhere, shall, underneath, unless, whence,* and *whither.* Although words such as *unless, nearby,* and *toward* seem to deserve inclusion, the general impression can be that the ISETTs or IETS? fare fairly well in presenting function words— something which can be attributed to the form-focused tradition in Iranian language teaching.

There are differences between the rankings of the top words in Table 2 and those in established references, i.e., Brown Corpus: *the, of, and, to, a, in*;

Cobuild General Corpus: *the, of, and, to, a, in*; the BNC: *the, of, and, a, in, to*; the GSL: *the, be, of, and, a, to*. The more frequent use of *be, a* and personal and demonstrative pronouns in IETSs can be due to the frequent use of demonstrative language in these textbooks, especially, in earlier stages, e.g., *This is a ….*, *She is a ….* The different line-up of the function words tells much about authenticity and representativeness of texts used.

Table 3 shows the content word families which feature in the top 200 words families from the full corpus. The word *lesson* is on the top due to its metatextual function and ranks 19stafter 18 function words. The end of the table displays examples of words in the next 100th intervals and examples of words with only three and two occurrences.

Table 3

The 125 Most Frequent Content Word Families in IETSs

lesson (19th)524	child 211	listen 142	hard 104	today 81	substitute 67
go 475	student 211	come 136	put 103	drive 77	feel 66
word 390	know 208	thing 130	learn 100	long 77	money 66
sentence 387	English 203	play (99th)129	talk 100	well 77	fast 64
use 353	time 201	year (101th)125	change 99	small 76	hand 64
answer 344	man 198	please 122	compare 96	wash 76	leave 64
teach 337	Ali 195	television 120	repeat 96	bicycle 75	letter 64
question 325	complete 182	take 119	usual 96	brother 75	sister 64
speak 307	day 179	want 119	Reza 95	Mynah 75	dialogue 63
write 294	like 177	boy 117	table 95	easy 73	homework (199th)63
follow 293	get 176	buy 117	partner 93	clock 72	….
book 286	yesterday 174	study 116	structure 93	country 71	farm(300th)42
school 284	new 167	home 115	find 92	late 71	…….
make 280	give 161	last 114	mother 92	door 70	tape(400th)30
work 260	help 159	house 111	verb 92	foot 70	comment(500th)22
read 258	good 157	ask 110	bus 91	need 70	idea(600th)17
example 256	model 157	practice 110	clean 90	past 70	solve(700th)14
very 237	watch 155	old 107	father 90	tell 70	fruit(800th)11
people 231	friend 153	say 107	pattern 88	week 70	London(900th)9
see 230	now 153	care 106	think 88	present 69	describe(1000th)7
car 225	eat 147		exercise 86	bed 68	addict(1370th)3
look 225	room 147		Mr. 84	night 68	advice(1521th)2
picture 224	live 145		pen 84	water 68	value(1521th)2
	morning 143		compute 82	moon 67	

Advice, value, Venus, vessel, wet, and *within* (a function word) are among the words with a frequency of two. *Account, adult, aerobics aftershock, and agent* are among words which occur once in the corpus.

Although all the words in Table 3 are very frequent in English, a cursory look at the table reveals a preponderance of metatextual words, especially those at the top of the list, e.g., *lesson, word, answer, and repeat.* To provide some clues to the representativeness of IETS texts and vocabulary, the ratings of some high frequency words were compared with their ratings in the GSL. The following juxtapositions may give a taste of how the frequencies of these words are compare with their frequencies in general English texts.

	lesson	*go*	*word*	*sentence*	*work*	*people*	*man*	*year*	home-work	*walk*
ISETTs	19	21	25	26	44	53	66	101	200	205
GSL	1473	50	178	1357	71	97	43	58	2276	341

In order to have an idea of the frequency of verbal chunks and the extent to which the books expose the learners to the associations among words, n-grams of 3-6 words were targeted, using AntConc program. Table 4 presents some of the most frequent n-grams in the full bank of IETSs, selected from the first 100 n-grams. The source list of n-grams in the output included more items, but only the more meaningful chunks were selected for presentation here. Obviously, most of the frequent n-grams are metatextual words from the instructions in the exercises.

Table 4
The 60 Most Frequent 3-6 Word Meaningful N-Grams in IETSs

look at the 125	write it down 47	the words in the pattern sentences 33
look at the pictures and 86	and complete the 46	there is a 33
with a partner 83	i don t 46	words in the pattern sentences 33
with a partner s 79	what time is 46	complete the following 32
answer these questions 76	the words in the pattern 44	the following words 32
answers with a partner 76	words in the pattern 44	what do you 30
compare your answers with a partners 76	substitute the words in the pattern 43	at the pictures and answer 29
a lot of 73	what time is it 43	at the pictures and complete 29
the words in 68	is it a 42	go to the 29
words in the 65	follow the example 38	look at the pictures and answer 29
go to school 59	listen to the 38	look at the pictures and complete 29
answer the questions 56	in the blanks 37	
follow the model 55	in the pattern sentences 37	in the picture 28
it is a 55	and answer the 36	new words and 28
complete the sentences 52	in the classroom 36	fill in the blanks with 27
on the table 50	in the morning 36	go to bed 27
listen and repeat 49	and answer the questions 35	in the park 27
substitute the words 49	fill in the blanks 35	new words and expressions 27
in the pattern 48	the following questions 34	
the words in the 48	substitute the words in the pattern sentences 33	
it s a 47		
the following sentences 47		

The most frequent n-grams are in the metatext of the books. In fact, the 100th n-gram in the full corpus has a frequency of 27, while the 100th n-gram in the subcorpus without metatext has a frequency of 13. To know about words associations in the body of lessons and exercises, the bank of texts without metawords was fed into AntConc program for n-gram analysis. Table 5 presents the more meaningful n-grams with frequencies above 20.

Table 5
N-Grams with Frequencies above 20 in IETSs without Metatext

a lot of 68	don't know 26
go to school 50	go to bed 25
it is a 50	in the park 25
on the table 43	do you know 23
it's a 42	I have a 23
what time is 42	in the evening 23
I don't 41	what did the 22
is it a 40	do you see 21
what time is it 39	it isn't 21
in the morning 35	of the moon 21
in the classroom 33	we don't 21
there is a 31	what is it 21
what do you 29	he has a 20
go to the 28	in the street 20
in the picture 28	

It is obvious that these books are not rich in presenting frequent English collocations and fixed phrases beyond the conventionalized language of instructions and a small number of basic constructions. A minority of structures is frequently recycled, but essential English phrases and constructions are absent or infrequent.

As indicated earlier, metalinguistic words play a significant role in language teaching books. They include words used in instructions and grammar boxes. Table 6 shows that words with frequencies above 2 are exclusively featuring in the instructions of these books. These words are based on a list subtracted from the full corpus. From words in Table 3, only *verb* and *substitute* feature here.

Table 6
The Metalinguistic Word Families and Key Morphemes with Frequencies
Higher Than Two

verb 92	particle 8	er 3
substitute 67	possess 7	identify 3
oral 50	bracket 6	ly 3
drill 43	ing 6	modals 3
false 35	position 6	section 3
pronunciation 31	affirmative 5	singular 3
comprehend 29	imagine 5	spell 3
parenthesis 29	precede 5	advisability 2
pronoun 22	previous 5	colon 2
passive 20	auxiliary 4	consequence 2
adverb 16	bare 4	contain 2
syllable 14	contraction 4	cue 2
gerund 13	item 4	determiners 2
phrase 13	modify 4	differ 2
tense 13	rela 4	error 2
preposition 12	stress 4	imperative 2
underline 12	wh 4	manner 2
participle 10	appear 3	omit 2
focus 9	appropriate 3	plural 2
vowel 9	comma 3	pron 2
count 8	either 3	pronounce 2
digest 8		

Table 7 displays a comparison of words in IETSs and those in the GSL, Ogden 850 basic words, the AWL, and Longman Communication 3000 to show how representative of the basic English vocabulary the IETSs are. IETSs share 1268 families with the GSL. The words unique to the series include proper nouns, mostly Iranian person and place names and partial or half words. Six-hundred and seventy families do not appear in the books at all. Of these, 154 families belong to the first 1000 GSL words and 516 families to the second

1000. One hundred and eighty-three Ogden words, 411 AWL words, and 1019 Longman words are missing from the books.

Table 7
IETS Vocabulary Compared with Some Well-Known Word Lists

Reference Lists / Levels	GSL 2000 Words			Ogden 850 Words			AWL			Longman 3000		
	Unique to GSL	Shared	Unique to IETS	Unique to Ogden	Shared	Unique to IETS	Unique to AWL	Shared	Unique to IETS	Unique to Longman	Shared	Unique to IETS
G1	1784	154	46	751	96	104	565	5	195	2336	147	53
G2	1525	313	119	650	197	235	561	9	423	2182	301	131
G3	1453	485	193	547	300	378	557	13	665	1990	493	185
H1	1207	731	257	430	417	571	539	31	957	1732	751	237
H2	1117	821	297	383	464	654	543	27	1091	1646	837	281
H3	1132	806	365	386	461	710	506	64	1107	1600	883	288
Pre-Uni	1028	910	503	341	506	907	444	126	1287	1410	1073	340
Books amassed	670	1268	1092	183	664	1696	411	159	2201	1019	1464	896
G1 Without Metatext	1840	98	28	786	61	65	569	1	125	2393	90	36
G2 Without Metatext	1679	259	83	876	171	171	566	4	338	2237	246	96
G3 without Metatext	1507	431	147	576	271	307	566	4	573	2053	430	148
H1 without metatext	1207	731	257	430	417	571	539	31	957	1732	571	237
H2 Without Metatext	1166	772	236	398	449	559	557	13	995	1710	773	235
H3 Without Metatext	1171	767	273	405	442	398	524	46	994	1652	831	209
Pre without metatext	1050	888	450	348	499	839	455	115	1223	1440	1043	295
Books amassed without metatext	698	1242	892	189	658	1476	433	137	1997	1061	1422	712

Table 7 shows that the series increasingly includes more words from these lists as it progresses up to the final level, but it fails to cover an overwhelming majority of the basic vocabulary items from these lists. Even in the case of Ogden 850 Basic Words, 189 words are left out. Only 137 AWL words are covered by these books. Although it seems pedagogically sound that only a few AWL or other content words are included in basic levels and the focus is on very basic words, the opportunities are not fully used later. A conspicuous example is H3, where only 3 new words are added from Ogden list. This is confirmed by checking the progress with other word lists in other columns. Moreover, it is a pressing need of the students at pre-university level to be familiar with academic words. The same number of AWL words appear in G2 and G3 and the number of AWL words in H2 is even lower than H1. The series ignores 433 AWL items.

The final piece of information reported here is the range of the words across the seven IETSs, i.e., how many of them repeat how many words. While it is acknowledged that this tabulation is very rough and short of a fine-tuned display of recycling of particular words in individual books and across books, combined with the frequency profiles above, it can be very telling. Table 8 indicates how many types and families of words are featured at each possible range, both in the main corpus and in the subcorpus excluding metatextual words.

Table 8
The Range of the Words across the IETSs

Range	1	2	3	4	5	6	7
Full-corpus Types	2028	662	393	306	232	180	134
Full-corpus Families	840	433	286	243	194	143	103
Subcorpus Types	1933	614	362	282	225	162	75
Subcorpus Families	839	415	275	226	191	129	55

A striking point in Table 8 is the high proportion of words occurring only in one book. The statistics for higher ranges are increasingly smaller. In fact, only 103 word families occur in all the seven books and only l43 families occur in six books.

Discussion

A brief look at the tables above reveals that IETSs fail to target an adequate number of English words and provide enough verbal exposure and practice for the words which are targeted.

According to Table 1, there are 1277 word families in the Pre-university book and a total of 1877 word families in all the seven books. Obviously, this seriously falls short of the minimum number of word families suggested by many researchers (e.g., Zahar, Cobb, & Spada, 2001; Laufer & Ravenhorst-Kalovski, 2010). In fact, IETSs do not go very much beyond half of the minimum number, i.e., 3000 words, which Laufer (1997) and Zahar, Cobb, and Spada (2001) find necessary for a functional understanding of the basic English texts. Table 7 shows that 698 GSL word families, 189 Ogden word families, 433AWL word families, and 1061 Longman 3000 word families are absent from the IETSs. This means that even in the ideal situation, where the targeted words were fully learned, high school graduates, who are mostly university candidates and aspire to study English at a higher level, would not be well prepared to handle either basic or academic English texts only for a lack in the vocabulary component. Thus, this fact about the vocabulary deficit of IETSs can be very helpful in diagnosing the problems university students face in learning English because the missing words in IETSs are very frequent in both academic and general texts.

Another issue is the insufficient recycling and practice of the targeted words. The frequencies presented along the words in most of the tables above and the range information in Table 8 bear witness to this fact. According to Table 3, 1000 words are repeated more than seven times in these books. Given the fact that there are 1877 words in IETSs, 1000 may sound like a good majority. However, it should be noted that a large number of these words are function or metatextual words. Metatextual words tend to become marginalized and not receive much attention not least because they are not in the main

passages of the lessons and are less likely to feature in exams. Thus, as a rough estimate, it seems that about half of the content words are not repeated more than seven times in these books. Unfortunately, we do not have a clear picture of the repetitions that do occur, as we do not know about the repetitions which are in the same lesson or passage or across multiple lessons in the same book. The only source of information about IETSs word recycling available here is Table 8, which gives the range of words across the seven books. The figures displayed in Table 8 again confirm the observation that there have not been serious systematic attempts at recycling the words which are introduced. On the one hand, there are a high proportion of words occurring in one book. On the other hand, only 103 word families occur in all the 7 books and only 143 families occur in six books. When the metatextaul words are subtracted from these words, only 55 and 129 word families are left with ranges of seven and six, respectively. Obviously, many of these high-ranging words are basic grammatical words. As a follow-up to the range analysis, the lists of words obtained from this analysis were compared with the Academic Word List. The comparisons revealed that only one AWL word has a range of seven, one a range of six, six occur in five books, four appear in four books, and 12 appear in three books.

Another dimension of the quantitative inadequacy in IETSs is the fact that a small number of familiar multiword fixed phrases or n-grams are included in IETSs (see Tables 4 & 5). Most of the frequent n-grams are in the metatextual parts of the books. In fact, there are only 29 English lexical chunks which occur more than 20 times in the main body of the lessons in IETSs. This is in spite of the emphasis put by experts on fixed phrases and collocations in teaching language. For example, Lewis (2000) vehemently advocates the inclusion and practice of lexical chunks in language teaching materials and Martinez and Schmitt (2012) cite research-based evidence which shows the importance of formulaic language.

The limited coverage of words, the low amount of recycling across the books and within individual books, the presence of few well-known English chunks and collocations, and lack of correspondence between the frequency ranking of the most frequent function words and their ranking in established reference corpora (see Table 2) indicate that the amount of text IETSs include

is not very much. This flies in the face of SLA research which emphasizes the quantity and frequency of exposure. Both those who support incidental exposure (e.g., Schmitt, 2008) and those who advocate direct instruction (e.g., Krashen, 2012) would agree that the inadequacy of input and exposure constitutes a threat to the vocabulary teaching, and ultimately the language teaching mission of IETSs.

Implications

The fairly detailed revelations of this study about the quantity of vocabulary input have noteworthy implications for policy makers, material developers, and teachers. It seems that seven years of English study deserve more lexical and textual input. It is very unlikely that such a grand goal as learning the functional vocabulary of a foreign language can be achieved by reading a small number of short passages. Insufficient input and exposure may be counterproductive not only for cognitive reasons, but also on affective grounds because such a treatment sends wrong messages about the effort needed for language learning and creates false expectations about success. But, before the yawning lexical and textual gaps are bridged by decision makers and material developers, teachers can draw upon the feedback from this study and prepare remedial supplementary vocabulary materials, using established pedagogical vocabulary references such as the AWL, the GSL, Longman Communication 3000, Ogden 850 Basic Words, and Oxford 3000. The bottom line of this study is that Iranian school students do need more exposure to basic and academic vocabulary as well as conventionalized phrases. This need can be met by taking more texts to the learners. The question of the method for doing this is another serious issue which is not the scope of this report.

Conclusion

Some areas of vocabulary input quantity are still in the dark. For example, the range of words across lessons in individual books could help teachers in setting up remedial practice and book developers in their later revisions. A list of words common to all or common to different combinations of the seven books could also be helpful. A comparative-correlational study of the frequency rankings of words in lists such as the GSL and the BNC and the frequency

rankings of the words in present textbooks would give further insight about the representativeness and priorities of these books concerning English vocabulary.

Nevertheless, the study provides a glimpse of the vocabulary that the last revision of IETSs attempts to teach. This information can be useful to teachers, book designers, and learners. As the report above shows, a considerable amount of basic and valuable English vocabulary is introduced and practiced. But, again, there is much, arguably equally important, vocabulary which is left out of these books. This may be due to lack of variety in text types or simply due to low input quantity or both. The books are not rich in idiomatic expressions and few frequent conventional English phrases make an appearance in them; therefore, they fail to create "an English atmosphere" for the users and give a taste of real English. This study did not directly target the amount of recycling and reuse of words within individual books. However, considering the lexical variation and range indices here, it is safe to say that, beyond function words and some intertextual items, many of the current words are not systematically and frequently revisited.

More in-depth studies can complement this study. For example, it is worthwhile to find if the current presentation order of the words is optimal or some reshuffling is required to improve their teachability. Research is needed to know how much recycling the average Iranian language learner needs to learn the targeted items. A fine-tuned study can identify the words which require more recycling, due to, say, complexity or being forgetting-prone. Another issue worthy of exploration is the extent to which the vocabulary deficiencies of undergraduate learners reflect the missing words in these books.

Notes on the Contributor:

Hadi Farjami is assistant professor at the Department of English Language and Literature, Semnan University, Iran. He has taught content and skills to students of English Language and Literature, Translation Studies, and Applied Linguistics for 18 years. He has also taught EAP courses to students of science and technology. He has published in international and national journals and authored and co-authored EFL textbooks.

References

Anthony, L. (2012). *AntConc (version 3.3.5w)*. Tokyo: Waseda University. Retrieved from http://www.antlab.sci.waseda.ac.jp/

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, *20*(2), 136-163.

Cobb, T. (2011). *The compleat lexical tutor*. Retrieved from http://www.lextutor.ca.

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

DeKeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125–151). New York: Cambridge University Press.

Flowerdew, J. (2001). Concordancing as a tool in course design. In M. Ghadessy, A. Henry, & R. L. Roseberry (Eds.). *Small corpus study and ELT* (pp. 71-92). Amsterdam: John Benjamin.

Ganger, J., & Brent, M. (2004). Reexamining the vocabulary spurt. *Developmental psychology*, *40*(4), 621-632.

Heather, J., & Helt, M. (2012). Evaluating corpus literacy training for pre-service language teachers: Six case studies. *International Journal of Technology and Teacher Education*, *20*(4), 415-440.

Herman, P., Anderson, R.C., Pearson, P.D., & Nagy, W. (l987). Incidental acquisition of word meaning from expositions with varied text features. *Reading Research Quarterly*, *22*, 263-274.

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *The Canadian Journal of Language Review*, *61*(3), 355–382.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition, 21*, 181-193.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Keshavarz, M. H., & Mohammadi, V. (2009). The effect of unknown vocabulary density on EFL learners' reading comprehension of nonfiction general English texts. *Journal of English Language Studies*, *1*(1), 1-22.

Krashen, S. (1989). We acquire vocabulary and spelling via reading: Additional evidence for the input hypothesis. *Modern Language Journal*, *73*(4), 440-464.

Krashen, S. D. (1993a). *The Power of Reading: Insights from the Research*. Englewood, CO: Libraries Unlimited.

Krashen, S. (1993b). The case for free voluntary reading. *Canadian Modern Language Review, 50*(1), 72-82.

Krashen, S. (2012). Reading and vocabulary acquisition: Supporting evidence and some objections. *Iranian Journal of Language Teaching Research*, *1*(1), 27-43.

Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady, & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 20–34). Cambridge: Cambridge University Press.

Laufer. B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*(1), 15-30.

Laufer, B., & Rozovski-Roitblat, B. (2011). Incidental vocabulary acquisition: The effects of task type, word occurrence and their combination. *Language Teaching Research*, *15*(4), 391-411.

Lessard-Clouston, M. (2013). *Teaching vocabulary*. Alexandria, VA: TESOL International Association.

Lewis, M. (2000). *Teaching collocations*. London: Language Teaching Publications.

Martinez, R., & Schmitt, N. (2012). A phrasal expression list. *Applied Linguistics, 33*(3), 299-320.

Milton, J. (2009). *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.

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

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

Nation, P. (2006). How large a vocabulary is needed for listening and reading? *Canadian Modern Language Review*, *63*(1), 59-82.

Nation, P., & Wang, K. (1999). Graded readers and vocabulary. *Reading in a Foreign Language*, *12*(2), 355-380.

Prichard, C., & Matsumoto, Y. (2011). The effect of lexical coverage and dictionary use on L2 reading comprehension. *The Reading Matrix*, *11*(3), 207-225.

Quirks, R. (1985). *A comprehensive grammar of English*. Harlow: Longman.

Reznick, J. S., & Goldfield, B. A. (1992). Rapid change in lexical development in comprehension and production. *Developmental Psychology*, *28*, 406-413.

Riazi, M., & Mosalanejad, N. (2010). Evaluation of learning objectives in Iranian high-school and pre-university English textbooks using bloom's taxonomy. *The Electronic Journal for English as a Second Language*, *13*(4). Retrieved from http://www.tesl-ej.org/wordpress/issues/volume13/ej52/ej52a5/

Richards, J. (2005). *The role of textbooks in a language program*. Cambridge: Cambridge University Press.

Roeper, T. (2007). What frequency can do and cannot do. In I. Gülzow, & N. Gagarina (Eds.), *Frequency effects in language acquisition* (pp. 23-50). Berlin: Mouton de Gruyter.

Römer, U. (2010). Using general and specialized corpora in English language teaching: Past, present and future. In M. C. Campoy-Cubillo, B. Belles-Fortuño, & L. Gea-Valor. (Eds.), *Corpus-based approaches to English language teaching* (pp.18-35). London: Continuum.

Rott, S. (2007). The effect of frequency of input-enhancements on word learning and text comprehension. *Language Learning*, *57*(2), 165-199.

Saragi, T., Nation, P., & Meister, G. F. (1978). Vocabulary learning and reading. *System*, *6*, 72-78.

Scholfield, P. (1991). *Vocabulary rate in coursebooks: Living with an unstable lexical economy*. Thessalonki: Aristotle University.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, *12*, 329-363.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sutarsyah, C., Nation, P., & Kennedy, G. (1994). How useful is EAP vocabulary for ESP? A corpus-based study. *RELC Journal*, *25*(2), 34-50.

Thorndike, E. L., & Lorge, I. (l944). *The teacher's word book of 30,000 words*. New York: Columbia University Press.

Walters, J. D. (2006). Methods of teaching inferring meaning from context. *RELC Journal, 37*(2), 176-190.

West, M. (1953). *A general service list of English words*. London: Longman.

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *The Canadian Modern Language Review*, *57*(4), 541-572.