An Evaluation of the Vocabulary Used in Iranian EFL Secondary and High School Textbooks Based on the BNC First Three 1000 High Frequency Word Lists

Mansour Koosha¹ - GholamReza Akbari²

Abstract
This study compares the vocabulary used in Iranian EFL secondary and high school textbooks with the first three 1000 high frequency word lists (1000, 2000, 3000) from the British National Corpus (BNC). The BNC first three word lists consist of 3000 word families which form the basis of 16454 word types. Using BNC lists and Nation's Range Program, the findings of this study show that while most of the words used in these textbooks are among the words most frequent used, they just make up 10% of the total words in the BNC first three 1000 high frequency word lists. This indicates that there is a

¹ Islamic Azad University, Khorasgan Branch, Mansour.Koosha@yahoo.com
² Islamic Azad University, Khorasgan Branch
wide gap between the size of the vocabulary knowledge presented in these textbooks and an ideal vocabulary knowledge (3000 word families) necessary for minimal reading comprehension. The results of this study help course designers and materials developers to reconsider the vocabulary selection in the preparation of language teaching texts and materials. They also inform teachers of the shortcomings in the textbooks and help them to try to find ways to bridge the gap.

**Key Words:** Vocabulary, Reading Comprehension, Word Type, Word Family, BNC Lists, Iranian EFL Secondary and High School Textbooks

1. Introduction

A large number of people around the world today use English as a second or foreign language. Many of these people learn English because their success in education depends on their ability to read the latest scientific and technical publications in English. Reading, therefore, is an important language skill that is now in more demand than at any time in our history. With the explosion of the information, students need to master reading, especially in order to understand the enormous amount of information presented through the Internet in the global arena. It is not surprising, therefore, that in many countries, where English is taught as a foreign language, as it is the case in Iran, the primary emphasis is placed on the development of reading comprehension ability. To achieve this goal, an adequate knowledge of words seems to be a major prerequisite. The role of vocabulary knowledge as an instrumental factor in reading comprehension has been confirmed in the field of reading and vocabulary research (Alderson, 2000; Laufer, 1996; Nation, 1990; 2001, 2004). As Doughty and Long (2003) state, large corpora of errors consistently show that most errors of second language learners in the process of communication are related to lexical errors. In the same vein, they cite a case study in which lexical errors, in comparison, were three times more than
grammatical errors. This difference gains special significance when we learn that whereas grammatical errors may result in erroneous structures which can be understood, lexical errors can seriously hamper the process of communication (Gass, 1988b).

In spite of the critical importance of vocabulary in learning a language, the field of vocabulary acquisition remained a neglected area until quite recently (Koosha & Sharifi, 2004). After a long period of being preoccupied with the development of grammatical competence, language teachers and second language acquisition (SLA) researchers now recognize the importance of vocabulary in language learning. The emphasis now placed on second language vocabulary acquisition research has resulted in the writing of several excellent books (Nation, 1999a, 2001, 2004; Read, 2000; Schmitt, 2000, Schmitt and McCarthy, 1997) and a growing number of research articles exploring the nature and the importance of vocabulary acquisition.

In reading assessment, vocabulary knowledge is viewed as an important factor. Alderson (2000) notes that “vocabulary plays a very important role in reading tests”, and that reading research has “consistently found a word knowledge factor on which vocabulary knowledge loads highly” (p. 99). Nation and Coady (1988) believe that “vocabulary knowledge would seem to be the most clearly identifiable subcomponent of the ability to read” (p. 98). Lewis (1993) asserts that “vocabulary (or lexis) carries more of the meaning of a text than does the grammar” (p. 17). Laufer’s research on the relationship between vocabulary size and reading comprehension has produced results indicating relatively high intercorrelations, ranging from .50 to .75, between the two factors.

An important issue in studies concerning the size of vocabulary required to read a text is the amount of text coverage, that is, the percentage of running words in the text known by the readers needed for adequate comprehension to occur. In other words, how much unknown vocabulary can be tolerated in a text by the readers? In this area, there have been several studies which attempt to determine the
amount of vocabulary needed by a language learner in order to be able to read with reasonable comprehension.

Hu and Nation (2000) examined the relationship between text coverage and reading comprehension for non-native speakers of English with a fiction text. The results showed that with a text coverage of 80% (that is, 20 unknown words out of every 100 words [1 in 5]), no one gained adequate comprehension. With a text coverage of 90%, a small minority gained adequate comprehension. With a text coverage of 95% (1 unknown word in 20), a few more gained adequate comprehension, but they were still a small minority. At 100% coverage, most gained adequate comprehension. When a regression model was applied to the data, a reasonable fit was found. It was calculated that 98% text coverage (1 unknown word in 50) would be needed for most learners to gain adequate comprehension. This figure fits with Carver’s (1994) findings with native speakers:

When the material being read is relatively easy, then close to 0% of the words will be unknown, ... when the material is relatively hard then around 2% or more of the words will be unknown, ... and when the difficulty level of the material is approximately equal to the ability level of the individual, then around 1% of the words will be unknown. (p. 432)

As Carver (1994) indicates, even 98% coverage does not make comprehension easy. Kurnia (2003), working with a non-fiction text, found that few L2 learners gained adequate comprehension with 98% coverage. These researchers believe that this extent of coverage requires more solid and rigorous investigation. The features of coverage cited above have led several researchers to consider whether there is a threshold level where vocabulary knowledge becomes sufficient for adequate comprehension. The assumption which lies behind this discussion is that vocabulary learning is strongly affected by word frequency. That is, words which occur frequently in the language tend to be learned before words that occur less frequently. There is some evidence to show that this is true. For example, Read
(1988) and Laufer, Elder, Hill, and Congdon (2004) found that learners’
scores dropped on the Vocabulary Levels Test [developed by Nation
(1990)] as students moved from higher to lower frequency levels. Laufer
and Nation (1999) found a similar effect for productive knowledge.

Research by Liu Na and Nation (1985) showed that this ratio of
unknown to known words is not sufficient to allow reasonably
successful guessing of the meaning of the unknown words. In their
research, low frequency words at two different densities in a text were
replaced by nonsense words. The subjects were asked to guess the
meanings of these words by using context clues. The results were then
analyzed in order to see what factors affected difficulty of guessing.
The results showed that words in the low density text (1 unknown
word in 25) were easier to guess than words in the high density text (3
words in 10). When the subjects were divided into proficiency levels
according to their success at guessing, learners at high proficiency
levels could successfully guess 85% to 100% of the unknown words.
The group of learners at the lowest proficiency level guessed between
30% and 40% of the unknown words.

In another research done by Hirsh and Nation (1992) on novels
written for teenagers or younger readers, the results indicated that
vocabulary size of 2000 to 3000 words provides a very good basis for
language use as reflected in Table 1.

<table>
<thead>
<tr>
<th>Vocabulary size</th>
<th>Coverage %</th>
<th>Density of unknown words</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000 words</td>
<td>90%</td>
<td>1 in every 10</td>
</tr>
<tr>
<td>2000 + proper nouns</td>
<td>93.7%</td>
<td>1 in every 16</td>
</tr>
<tr>
<td>2600 words</td>
<td>96%</td>
<td>1 in every 25</td>
</tr>
<tr>
<td>5000 words</td>
<td>98.5%</td>
<td>1 in every 67</td>
</tr>
</tbody>
</table>

*Table 1: Vocabulary size and coverage in novels for teenagers
(Hirsh and Nation, 1992)*
Coady (1997) proposed that good knowledge of at least 5000 word families in the L2, in addition to significant reading skill, is required for understanding academic texts. In the latest research conducted by Nation (2006) on the vocabulary size, it is asserted that if we take 98% as the ideal coverage, a 8000–9000 word-family vocabulary is needed for dealing with written text, and 6000–7000 families for dealing with spoken text. In this vein, using a current 5 million word spoken corpus, Adolphs and Schmitt (2004) found that 2000 word families supply lexical coverage for less than 95% of spoken discourse.

Therefore, according to the findings of the afore-mentioned studies, it can be concluded that the threshold level of vocabulary knowledge for adequate reading comprehension is around 95 percent of the words in a given text (Hu & Nation 2000; Laufer 1989, 1997; Liu & Nation 1985; Wixson & Lipson 1991). This means that basic reading comprehension may require readers to know at least 19 words out of every 20 words they encounter in a given text, and that the meaning of an unknown word is more easily understood from context if approximately 19 neighboring words are known. This is only possible when a minimum of 95 percent of the words in a text is already known to the reader. However, determining an exact cut-off point for the text coverage is a controversial issue. While researchers may reluctantly agree on the approximate number of words that constitute 95 percent of most texts, there appears to be less disagreement on the 95 percent comprehension threshold itself. According to Schmitt (2001) for the reading of an English text a threshold of 3000 word families is required for minimal reading comprehension. Laufer (1998) believes that knowing fewer than 3000 word families indicates poor L2 reading ability regardless of L1 reading ability.

On the other hand, since it is impossible to teach the whole vocabulary of a language, we should select those lexical items which are more useful for the language learner. One measure of usefulness is
word frequency, that is, how often the word occurs in the normal use of the language. Fortunately, in recent decades, extensive analysis of corpora has offered textbook writers, teachers and learners of English a solution to the problem of vocabulary selection by identifying lists of high frequency words that make up the core of the language (see for example, Nation 2006).

The distinction between high frequency and low frequency words is critical for language teachers and course designers. Second language (L2) vocabulary researchers believe that the learner needs to know the 3000 or so high frequency words of the language (Laufer, 1998; Schmitt, 2001). These are an immediate high priority and there is little sense in focusing on other vocabulary until these are well learned.

Lists of high frequency words provide an important basis for deciding what goes into language courses and supplementary materials. They also provide a trusted source for direct vocabulary learning. In this vein, the fourteen word family lists which were developed by Paul Nation from the British National Corpus (BNC) are the most recent frequency lists. Fourteen 1000 word family lists were made from the first 1000 to the fourteenth 1000. The BNC first three 1000 high frequency word lists consist of 3000 word families which form the basis of 16454 types. For Nation (2010), a type consists of a headword and its inflected forms where the headword and its inflected forms are all the same part of speech. Types are different word forms that are the sources of tokens. Tokens are occurrences of types, so one type can have many tokens i.e. it can occur many times. To count tokens, each word is counted one after another and even if the same word form occurs again it is counted again. It is the types that make word families. Word families include both closely related inflected and derived forms even if the part of speech is not the same (Nation, 2010, personal communication through e-mail).

The BNC first 1000 list consists of the first set of 1000 word families and these 1000 word families include 6348 types. The BNC
2nd 1000 list consists of the second set of 1000 word families which include 5592 types. Similarly, the BNC 3rd 1000 list contains the third set of 1000 word families that cover 4514 types.

Nation (1990) maintains that after these high frequency words are learned, the next focus for the teacher is on helping the learners develop strategies to comprehend and learn the low frequency words of the language. Because of the very poor coverage that low frequency words give, it is not worth spending class time on teaching these words. The significance of the information provided by the studies in the field of vocabulary and reading is that although there are over 54000 word families (Goulden, Nation & Read, 1990) in English, and educated adult native speakers know around 20000 of these word families (Goulden, Nation, & Read, 1990; Zechmeister & Chronis, 1995), a much smaller number of words, 3000 word families, is needed to provide a basis for reading comprehension.

School teaching programs can help learners to master at least the minimum number of words need for effective and efficient reading. ELT materials (textbooks) play a very important role in this connection. Sheldon (1988) suggests that "textbooks represent the visible heart of any ELT program" (p. 237). Cunningsworth (1995) contends that we should ensure "that careful selection is made, and that the materials selected closely reflect the needs of the learners and the aims, methods, and values of the teaching program" (p. 7).

In Iranian secondary and high schools, the English teaching program is composed of seven English textbooks designed by the Ministry of Education to meet the needs of students most of whom hope to major in different areas of humanities, science and technology in the future. The main objective of the English textbooks they study is to help them develop the skills of reading general, scientific and technical texts in English needed in many areas for furthering their education.
Due to the fact that textbooks are used directly as instructional materials and they are used as the main tools in the teaching and learning process, in this study the researchers investigated the selection of the vocabulary used in Iranian EFL secondary and high school textbooks to see if careful selection has been made and the vocabulary used closely reflects the needs of the learners and are in line with the latest findings in the area of vocabulary studies and the aims of the teaching program. This study, therefore, in order to compare the kind and size of the vocabulary used in Iranian EFL secondary and high school textbooks, by using the BNC lists and Nation’s Range Program (a computer software used to compare the word lists under study with the BNC word frequency lists), analyzed the word lists of the mentioned textbooks to see if the words of the highest frequency have been selected. The reason behind raising this issue is that in spite of the fact that although a language makes use of a large number of words, not all of these words are equally useful for all contexts and all purposes. Researchers in the field of materials development such as Tomlinson (2002) believe that materials developers and course designers should apply reliable sources and criteria for the preparation of language teaching texts and materials. According to the L2 vocabulary researchers one of the most applicable and reliable criteria for choosing words in preparing language teaching texts and materials is word frequency. In sum, as stated before, the study investigated the extent to which these textbooks present the threshold level of vocabulary knowledge required for minimal reading comprehension.

2. Materials
The materials used for the purposes of the present study were Iranian EFL secondary and high school textbooks, the first three BNC word frequency lists (1000, 2000, 3000 levels) and Nation’s Range program which are presented below in detail.
2.1. **Secondary School Textbooks**
1. The English textbooks of Iranian secondary schools which are presented below in detail.

2.2. **High School Textbooks**
2. The English textbooks of Iranian high schools which are presented below in detail.

2.3. **BNC word frequency lists**
The first three BNC word frequency lists (1000, 2000, 3000 levels) are shown in detail in Table 2.

*Table 2: Number of types and families in the first three BNC word frequency lists*

<table>
<thead>
<tr>
<th>BNC Word Lists</th>
<th>Number of Types</th>
<th>Number of Families</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>6348</td>
<td>1000</td>
</tr>
<tr>
<td>Two</td>
<td>5592</td>
<td>1000</td>
</tr>
<tr>
<td>Three</td>
<td>4514</td>
<td>1000</td>
</tr>
</tbody>
</table>
2.4. **Nation’s Range Program**: RANGE was designed by Paul Nation and Averil Coxhead of the School of Linguistics and Applied Language Studies, Victoria University, P.O. Box 600, Wellington, New Zealand. This program has been validated in a study conducted by Laufer and Nation (1995) and has been widely used in the domain of vocabulary studies. Since Range program is accompanied by special frequency-based word lists, it deconstructs any text or corpus into its lexical components by their frequency zones through the following procedures: It takes a given text as the raw input (the text may be typed, pasted or scanned into the program); checks the lexis of that text against its accompanying frequency-based word lists; and finally, as output, generates a lexical frequency profile (LFP) of that text in just a few seconds. The LFP generated as such describes the lexical content of a text in terms of four frequency zones which are actually representative of the four word lists ordinarily available in the program:

The first 1000 most frequent words in General English (GE); the second 1000 most frequent words in GE (i.e. from 1001 to 2000); the third 1000 most frequent words in GE (i.e. from 2001 to 3000) and; the words not included in any of the above lists (NIL or ‘not in the Lists) so they are normally addressed as the ‘difficult’ or low frequency words.

The word lists of Iranian EFL secondary and high school textbooks were extracted from the textbooks and changed into text documents in order to be run through Nation’s Range Program in several stages. At the first stage, the word lists in EFL secondary textbooks were run through Nation’s Range Program to be analyzed. In the second stage, the word lists of EFL high school textbooks were run through Nation’s Range Program in order, i.e. from book one to four. And finally both word lists, i.e. the word lists of Iranian EFL secondary and high school textbooks were run through Nation’s Range Program and the data obtained from the analyses were arranged into separate tables in the form of numbers and percentages.
3. Results
Concerning the first research question; that is, whether the lexical items used in Iranian EFL secondary and high school textbooks are in line with the first three BNC word frequency lists, the word lists of Iranian EFL secondary and high school textbooks were run through Nation’s Range Program in the following order.

At the first stage, the word lists in all three EFL secondary textbooks were run through Nation’s Range Program and the following results were found which are summarized in Table 3.

Table 3: Number and percentage of the word types and word families of the word lists in English books of secondary schools found at each of the first three British National Corpus (BNC) word lists

<table>
<thead>
<tr>
<th>Word List</th>
<th>Word Types</th>
<th>Word Families</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percentage in the textbook</td>
</tr>
<tr>
<td>List One</td>
<td>439</td>
<td>76.61</td>
</tr>
<tr>
<td>List Two</td>
<td>77</td>
<td>13.44</td>
</tr>
<tr>
<td>List Three</td>
<td>19</td>
<td>3.32</td>
</tr>
<tr>
<td>Not in the lists</td>
<td>38</td>
<td>6.63</td>
</tr>
<tr>
<td>Total</td>
<td>573</td>
<td>100</td>
</tr>
</tbody>
</table>

As Table 3 shows, 439 word types in the textbook word lists are in the base List One and these 439 word types make up 76.61% of the total word types in the textbook word lists and 2.66% of the total
word types in the BNC first three base word lists. The second row reveals that 77 word types in the textbook word lists are in the base List Two and these 77 word types make up 13.44% of the total word types in the textbook word lists and 0.46% of the total word types in the BNC first three base word lists. The third row indicates that 19 word types in the textbook word lists are in the base List Three and these 19 word types make up 3.32% of the total word types in the textbook word lists and 0.12% of the total word types in the BNC first three base word lists. And finally 38 word types which make up 6.63% of the total word types in the textbook word lists were not found in the first three base word lists.

In the second stage, the word list in English Book One (EB1) of high school was run through Nation’s Range Program and the following results were found which are summarized in Table 4.

Table 4: Number and percentage of the word types and word families of the word list in English Book One (EB1) of high school found at each of the first three British National Corpus (BNC) word lists

<table>
<thead>
<tr>
<th>Word List</th>
<th>Word Types</th>
<th>Word Families</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percentage in the Text</td>
</tr>
<tr>
<td>One</td>
<td>320</td>
<td>67.94</td>
</tr>
<tr>
<td>Two</td>
<td>87</td>
<td>18.47</td>
</tr>
<tr>
<td>Three</td>
<td>24</td>
<td>5.10</td>
</tr>
<tr>
<td>Not in the lists</td>
<td>40</td>
<td>8.49</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>471</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>
As Table 4 indicates, there are 471 word types in the word list of EB1 from which 431 word types are in the first three BNC word lists and 40 word types which make up 8.49% of the total word types in the textbook word list were not found in the first three base word lists. These 431 word types make up 2.62% of the total word types in the BNC first three base word lists. Table 4 also reveals that there are 380 word families in the textbook word list which make up just 12.65% of the total word families in the BNC first three base word lists.

In the third stage, the word list in English Book Two (EB2) of high school was run through Nation’s Range Program and the following results were found which are summarized in Table 5.

### Table 5: Number and percentage of the word types and word families of the word list in English Book Two (EB2) of high school found at each of the first three British National Corpus (BNC) word lists

| Word List | Word Types | | | Word Families | | |
|-----------|------------|----------------|----------------|----------------|----------------|
|           | Number     | Percentage in the Text book | Percentage in the first three BNC word lists | Number | Percentage in the Text book | Percentage in the first three BNC word lists |
| List One  | 217        | 57.87          | 1.32       | 197 | 62.34 | 6.56         |
| List Two  | 89         | 23.73          | 0.54       | 85 | 26.90 | 2.83         |
| List Three| 35         | 9.33           | 0.22       | 34 | 10.76 | 1.13         |
| Not in the lists | 34 | 9.07 | - | - | - | - |
| **Total** | **375**    | **100**        | **2.08**   | **316** | **100** | **10.52**   |

As Table 5 reveals, out of 375 word types 217 word types in the textbook word list are among the most frequent word types. Moreover,
34 word types which make up 9.07% of the total word types in the textbook word list were not found in any of the first three base word lists. In addition, it can be seen that there are 316 word families in the textbook word list which make up just 10.52% of the total word families in the BNC first three base word lists.

In the fourth stage, the word list in English Book Three (EB3) of high school was run through Nation’s Range Program and the following results were found which are summarized in Table 6.

Table 6: Number and percentage of the word types and word families of the word list in English Book Three (EB3) of high school found at each of the first three British National Corpus (BNC) word lists

| Word List | Word Types |  | Word Families |  |
|-----------|------------|-------------------------------|-------------------------------|
|           | Number     | Percentage in the Text book   | Percentage in the first three BNC word lists | Number | Percentage in the text book | Percentage in the first three BNC word lists |
| List One  | 218        | 59.73                         | 1.33                          | 202     | 64.10                       | 6.73                          |
| List Two  | 89         | 24.38                         | 0.54                          | 84      | 26.70                       | 2.80                          |
| List Three| 32         | 8.77                          | 0.20                          | 29      | 9.20                        | 0.97                          |
| Not in the lists | 26         | 7.12                          | -                             | -       | -                           | -                             |
| Total     | 365        | 100                           | 2.07                          | 315     | 100                         | 10.50                         |

As Table 6 shows, from 365 word types, 218 word types in the textbook word list are in the base List One which make up 59.73% of the total word types in the textbook word lists and 1.33% of the total word types in the BNC first three base word lists. In addition, 26 word...
types which make up 7.12\% of the total word types in the textbook word list were not found in the first three base word lists.

Table 6 also shows that from 3000 high frequency word families, 315 word families were found in the word list of the textbook under study which make up 10.50\% of the total word families in the BNC first three base word lists.

In the fifth stage, the word list of English Book Four (EB4) of high school was run through Nation’s Range Program and the following results were found which are summarized in Table 7.

**Table 7: Numbers and percentages of the word types and word families of the Word List in English Book Four (EB4) of high school found at each of the first three British National Corpus (BNC) word lists**

| Word List | Word Types | | Word Families | |
|-----------|------------|---|---|---|---|---|
|           | Number     | Percentage in the Text book | Percentage in the first three BNC word lists | Number | Percentage in the Text book | Percentage in the first three BNC word lists |
| List One  | 174        | 37.50 | 1.06 | 159 | 49.69 | 5.30 |
| List Two  | 118        | 25.43 | 0.72 | 113 | 35.31 | 3.77 |
| List Three| 49         | 10.56 | 0.30 | 48  | 15   | 0.50 |
| Not in the lists | 123 | 26.51 | - | - | - | - |
| Total     | 464        | 100  | 2.08 | 320 | 100  | 9.57 |

As Table 7 reveals, out of 464 word types, just 174 word types making up 37.5\% of the total word types in the textbook word list are in the base List One. Interestingly, 123 word types which make up 26.51\% of the total word types in the textbook word list were not found...
in any of the first three base word lists. In addition, it can be seen that there are 320 word families in the textbook word list making up just 9.57% of the total word families in the BNC first three base word lists.

In the sixth stage, the word lists in four Iranian EFL high school textbooks were run through Nation’s Range Program and the following results were found which are summarized in Table 8.

**Table 8:** Number and percentage of the word types and word families of the word lists in four Iranian EFL high school textbooks found at each of the first three British National Corpus (BNC) word lists

<table>
<thead>
<tr>
<th>Word List</th>
<th>Word Types</th>
<th></th>
<th></th>
<th></th>
<th>Word Families</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percentage in the Text book</td>
<td>Percentage in the first three BNC word lists</td>
<td>Number</td>
<td>Percentage in the text book</td>
<td>Percentage in the first three BNC word lists</td>
<td></td>
</tr>
<tr>
<td>List One</td>
<td>740</td>
<td>50.62</td>
<td>4.50</td>
<td>599</td>
<td>56.77</td>
<td>19.97</td>
<td></td>
</tr>
<tr>
<td>List Two</td>
<td>364</td>
<td>24.90</td>
<td>2.22</td>
<td>325</td>
<td>30.81</td>
<td>10.83</td>
<td></td>
</tr>
<tr>
<td>List Three</td>
<td>138</td>
<td>9.44</td>
<td>0.84</td>
<td>131</td>
<td>12.41</td>
<td>4.37</td>
<td></td>
</tr>
<tr>
<td>Not in the lists</td>
<td>220</td>
<td>15.05</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1462</td>
<td>100</td>
<td>7.56</td>
<td>1055</td>
<td>100</td>
<td>35.17</td>
<td></td>
</tr>
</tbody>
</table>

Table 8 indicates that, from 1462 word types, 740 word types in the textbook word lists are in the base List One. These 740 word types make up 50.62% of the total word types in the textbook word lists. It indicates that about 50% of the total word types and 56% of the word families in the word lists of Iranian EFL high school textbooks are found in the base List One. These 740 word types make up just 4.50% of the total word types in the BNC first three base word lists and 599
word families make up just 19.97% of the total word families in the BNC first three base word lists. Moreover, 220 word types which make up 15.05% of the total word types in the textbook word lists were not found in any of the first three base word lists. Finally, it can also be seen that there are 1462 word types and 1055 word families in the textbook word lists which make up just 7.56% of the total word types in the BNC first three base word lists and 35.17% of the total word families in the BNC first three base word lists. And finally concerning the second research question, i.e., whether the Iranian EFL secondary and high school textbooks present the threshold level of vocabulary knowledge required for minimal reading comprehension, the word lists in three Iranian EFL secondary textbooks and four EFL high school textbooks were run through Nation’s Range Program and the following results were found which are summarized in Table 9.

**Table 9: Number and percentage of the word types and word families of the word lists in four Iranian EFL high school textbooks and three EFL secondary textbooks found at each of the first three BNC word lists**

<table>
<thead>
<tr>
<th>Word List</th>
<th>Word Types</th>
<th>Word Families</th>
</tr>
</thead>
<tbody>
<tr>
<td>List One</td>
<td>1010</td>
<td>777</td>
</tr>
<tr>
<td>List Two</td>
<td>430</td>
<td>388</td>
</tr>
<tr>
<td>List Three</td>
<td>154</td>
<td>147</td>
</tr>
<tr>
<td>Not in the lists</td>
<td>257</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>1851</td>
<td>1312</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number in the Textbook</th>
<th>Percentage in the Textbook</th>
<th>Percentage in the first three BNC word lists</th>
<th>Number in the textbook</th>
<th>Percentage in the first three BNC word lists</th>
</tr>
</thead>
<tbody>
<tr>
<td>List One</td>
<td>54.57</td>
<td>6.14</td>
<td>59.22</td>
<td>25.90</td>
</tr>
<tr>
<td>List Two</td>
<td>23.23</td>
<td>2.62</td>
<td>29.57</td>
<td>12.93</td>
</tr>
<tr>
<td>List Three</td>
<td>8.32</td>
<td>0.94</td>
<td>11.21</td>
<td>4.90</td>
</tr>
<tr>
<td>Not in the lists</td>
<td>13.88</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>9.70</td>
<td>100</td>
<td>43.73</td>
</tr>
</tbody>
</table>
As Table 9. indicates, these textbooks present 1851 word types from which 1594 word types were found in the BNC first three word lists and 257 word types were outside the mentioned lists. These 1594 word types make up just about 10% of the total word types in the BNC first three word lists which include 16454 word types. In addition, the number of the word families existent in all seven textbooks under study is 1312 which make up 43.73% of the total word families in the BNC first three base word lists.

4. Discussion

The results obtained from the analysis of all three EFL secondary textbooks revealed that about 77 percent of the total word types in the textbook word lists were found in the first base word list which makes up about 2.66% of the total word types in the BNC first three 1000 high frequency word lists. Totally, about 93 percent of the total word types in the textbook word lists appeared in the BNC first three 1000 high frequency word lists which make up just about 3.24% of the total word types in the BNC first three 1000 high frequency word lists. On the other hand, if we consider word family as the basis of the comparison, this number of word families, i.e. 461 makes up just about 15.36 of the total word families in the BNC first three 1000 high frequency words.

And finally about 7 percent were outside the BNC first three 1000 high frequency word lists. Including this percentage of low frequency words at this level may not be justifiable. As L2 researchers in the field of vocabulary learning contend, in the early stages of learning a language, language learners should be exposed to high frequency words and learning low frequency words should be postponed to the later stages.

The results also showed that 38 words were not found in any of the BNC first three 1000 high frequency word lists. Among these 38 words, there were proper names such as names of people and places
which relate to the local contexts of the passages used and were only natural to be introduced. One has to remember that frequency of words is not the sole factor in the selection of vocabulary. Availability, learnability, coverage, frequency and range can sometimes compete with one another. Besides, it seems to be very difficult to write a natural sounding passage based only on the frequency factor.

Considering the findings obtained from the analysis of EB1 of high school, we can see that Iranian learners become familiar with 471 words during the first year of high school, which do not seem to expose the learners with sufficient exposure to the vast store of vocabulary needed in their future studies. Meanwhile, among these 471 words, 40 words are low frequency words which may not be easily justifiable at this level.

The results obtained from the analysis of EB2 of high school indicated that about 10% of the total words in EB2 is outside the high frequency words. In addition, if one compares EB2 with EB1 it can be seen that the percentage of the words of EB2 appearing in the first BNC word list is 57.87%, which is about 10% less than what is introduced in EB1. Furthermore, regarding the point that language learners are exposed to about 100 hours instruction during the second year of high school, the number of 375 words for a whole year is not sufficient. The difference in terms of sheer numbers of words between EB1 & EB2 is 96 words, that is, learners in the second year learn far fewer words than those in the first year.

Referring to the findings obtained from the analysis of EB3 of high school, we can see that learners in the third year who are more mature and have more familiarity with English structures are taught just 365 words. This number of words does not match with the amount of time, approximately, 100 hours, that EFL learners are exposed to language instruction during this year.
The results obtained from the analysis of EB4 of high school indicated that, as far as frequency is concerned, while about 37 percent of the total word types in the word list of this textbook are among the first base word list and about 36 percent exist in the second and third base lists, about 27 percent do not appear in the BNC first three 1000 high frequency word lists. Including this percentage of low frequency words may not be justifiable here because learners have not yet become familiar with the large number of the most frequently words appearing in the first 1000 high frequency word list. However, the authors may have their own reasons for their inclusion.

Another considerable issue which is also common to the other books of the same series, is that, based on the curriculum, these learners are exposed to about 150 hours of language instruction during the fourth year of high school. Therefore, they are expected to have become familiar with a far larger number of words due to their cognitive abilities and the amount of time available for language learning. Comparing this book with the previous books, it can be seen that a drastic change has occurred from the frequency point of view.

The results obtained from the analysis of all four English textbooks of high school indicated that about 50% of the word types in all four textbooks were among the BNC first 1000 high frequency word list, which makes up just about 4.50% of the total word types in the BNC first three 1000 high frequency word lists. About 35% appeared in the second and third base word lists which makes up just about 3.6% of total word types in the BNC first three 1000 high frequency word lists. And 15 percent were outside the BNC first three 1000 high frequency word lists. These results revealed that the authors of the book may not have considered word frequency as one of the main criteria in vocabulary selection.

In addition, findings of this analysis showed that there was a tendency toward using less frequent words in developing these textbooks. Including this percentage of low frequency words is not
justifiable here because Iranian students have not become familiar with a large number of the most frequent words appearing in the first 1000 high frequency word list. Up to this point, out of 6348 word types in the first base word list Iranian language learners have been exposed to just 1010 word types. This shows that there is a wide gap between the desirable number of frequent words students should be exposed to and the real number of frequent words students have been exposed to.

Therefore, it seems reasonable to familiarize the language learners with the rest of the most frequent words first which are about more than 5000 word types before dealing with the less frequent words. Another problem is that due to the language learners’ cognitive ability and the amount of time available for language learning presenting this number of words may not be justifiable.

And finally, the results obtained from the analysis of all seven English textbooks showed that about 86% of the total word types in the word lists of the mentioned textbooks were found in the BNC first three 1000 high frequency word lists. Of course, it should be noted that these 86% of words consist of 1594 word types appearing in the textbooks which were found among 16454 word types appearing in the BNC first three 1000 high frequency word lists. These 1594 word types make up just about 10% of the total word types in the BNC first three 1000 high frequency word lists.

The results also revealed that 257 word types were outside the lists of the most frequent words.

According to Nation (2001), the BNC lists were sorted by frequency and the most frequent words appear in the first list of BNC. Thus, the authors of the textbooks could have included more words from the most frequent ones in the textbooks provided the contexts were appropriate.

In a research done by Laufer (1998), it was found that over six years of instruction, learners’ receptive vocabulary size reached 1900
word families. By the end of the following year; however, it rocketed to 3500 word families. Similarly, productive vocabulary averaged 1700 word families after six years and rose to 2550 at the end of the following year. Waring (1997) found similar results. These studies indicate that a productive or receptive vocabulary size of around 2000 word families is a possible goal after 1000 hours of study of English as a foreign language.

Therefore, in language teaching programs in Iran where English is taught as a foreign language, learners study English about two to four hours per week for about 36 weeks a year during seven years—very approximately 1000 hours. In such a program we would expect learners to have become familiar with the high frequency words of the language and to have worked on strategies for dealing with low frequency words. If learners leave school without control of the high frequency words then their learning will have been for little purpose. However, the major problem that emerges is that among 16454 words existent in the BNC first three 1000 high frequency word lists, Iranian EFL learners become familiar with just 1594 words which make up around 10% of the total words in the BNC first three 1000 high frequency word lists while they are exposed to about 1000 hours of instruction during seven years of language learning in Iranian schools and this amount of vocabulary knowledge presented by these textbooks is not sufficient for the minimal reading comprehension and there is a wide gap between the amount of the vocabulary knowledge presented by the afore-mentioned textbooks (1312 word families) and a desirable vocabulary knowledge (3000 word families) which is necessary for minimal reading comprehension.

The majority of students learning a second or foreign language consider vocabulary as their major priority and agree that most of their difficulties in the receptive and productive use of language arise from their inadequate vocabulary knowledge (Nation, 1990). Iranian learners of English are aware of their limitations in recognition and use of
English words and perceive lack of vocabulary as their major problem. A good documentation for this is their poor performance on the vocabulary section of university entrance exams at different levels. The researchers’ experiences imply the fact that a large number of high school students confirm that inadequate vocabulary knowledge is their major problem in reading a text.

In sum, the results of this study in conjunction with the empirical researches conducted by Koosha (2001) and Sharifi (2004) on the lexical development and the vocabulary size of Iranian EFL learners revealed that the vocabulary knowledge of Iranian EFL learners before entering university and also in the first year of studying at university is far below the acceptable threshold level required for minimal reading comprehension and meaningful communication.

5. Conclusion
According to McCarthy (1990), “No matter how well the student learns grammar, no matter how successfully the sounds of L2 are mastered, without words to express a wide range of meanings, communication in an L2 just cannot happen in any meaningful way” (p. viii).

A large proportion of the world’s English learners are studying English in order to learn reading skill as the prominent language skill they need. Acquiring the lexical items of another language is a daunting task for language learners. But the task becomes more manageable if we know which words are more important to learn than others, or which words are most useful to know as a precondition to learning others. In English, computational studies of word frequency and text coverage, in conjunction with empirical studies of learner comprehension of texts with different lexical profiles, have provided valuable information for course designers, teachers and learners. It has become clear that words of high frequency are more important in the
early stages of learning a language. Summarizing the answers for the research questions, the researchers have found that the word lists in Iranian EFL secondary and high school textbooks are to a large extent in line with the BNC first three 1000 high frequency word lists but the vocabulary knowledge presented by the textbooks under study falls below an expectable threshold level required for minimal reading comprehension. One of the solutions to the problem is that the writers of the textbooks do a drastic revision based on frequency criterion in order to improve the content of the textbooks.

Moreover, it can be concluded that using the lists of high frequency words, English teachers can close the existent gap by presenting the lexical items which were not introduced in the textbooks. Of course, they should bear in mind that in spite of all advantages these frequency lists have, researchers such as White (1997) believe that there are a number of criteria that can be used in the selection and gradation of vocabulary. He argues that frequency (the total number of occurrences of an item in a given corpus of language), coverage (the number of things which can be expressed by any given item), range (the amount of times a word or words appear in texts within a given corpus), availability (the readiness with which a word is remembered and used by native speakers in certain situations), and potential learnability (the ease or difficulty with which a particular word can be acquired) can all play an important role in vocabulary selection. From the learner’s point of view, a crucial factor in L2 vocabulary acquisition regardless of word frequency, is word learnability. This is the ease or difficulty with which a particular word can be acquired. Two words may have the same frequency, but one may be more difficult to learn than the other due to factors which have to do with the features of the word, or with other words related to it in the target language, or in the learner’s L1 (Laufer 1990, 1997, Swan 1997). For example, a word which is a cognate in a learner’s L1 may be infrequent, but it may present no difficulty in learning. However,
teachers need to have reference lists to judge whether a particular word deserves attention or not, and whether a text is suitable for a class, and course designers should have lists to refer to when they consider the vocabulary as a component of a language course.

And finally, to teach vocabulary, it is necessary to recreate the contexts in which words are used in settings and situations which the learner is more likely to encounter. Exposing language learners to a wide variety of texts which are at the level of the readers’ linguistic level of proficiency, age, interests, future needs and aspirations, combined with the teaching of effective reading strategies used in both intensive and, especially in the case of vocabulary development, extensive reading can be a promising step forward. It is also worth mentioning the fact that noticing plays an important role in vocabulary development. According to Elgort (2007) deliberate teaching of vocabulary can result in both explicit and implicit knowledge of words.

References


