Discovering EFL learners’ perception of prior knowledge and its roles in reading comprehension

Zheng Lin

School of Education, University of Ballarat, Australia

This paper investigates EFL learners’ perception of prior knowledge and its roles in reading comprehension. It is based on a survey conducted among 400 EFL students studying at secondary and tertiary levels in China. Through the analysis of the ranked multiple responses to the questions posed in the questionnaire, the paper shows that EFL students in mainland China believe that their command of English vocabulary plays a crucial role in their reading comprehension. However as their linguistic knowledge increases, they tend to attach less importance to their linguistic knowledge, especially the knowledge of English syntax and formal structures. At the same time, conceptual and sociocultural knowledge seems to gain greater importance. Furthermore, the perceived importance of linguistic knowledge seems to start diminishing around the end of secondary education, when EFL students have acquired a vocabulary of about 3000 words and the basic knowledge of English syntax and formal structures. The final replacement of linguistic knowledge by conceptual or sociocultural knowledge as the top factor that affects their reading comprehension seems to take place one year after the beginning of the tertiary EFL course. This may well be considered as an indication of the EFL threshold level for EFL students in China.

It is widely accepted in the reading research community that readers’ prior knowledge plays an important role in reading comprehension (Adams, 1990; Carrell, 1983, 1984b, 1991; Grabe, 1991; Ulijn & Salager-Meyer, 1998; Weaver, 1994). Although current emphasis in reading research seems to focus on the use of prior knowledge in reading, it would help to have a better understanding of the use if we can have a clearer idea of what prior knowledge has been used, or what prior knowledge that readers think they have used in reading comprehension.

Eskey (1986) refers to the reader’s prior knowledge as ‘knowledge crucial to reading’, which is then categorised into two types: ‘knowledge of form’ and ‘knowledge of substance’ (p. 18). The knowledge of form provides the reader with certain expectations about the language of the text and facilitates making accurate identifications of forms in reading. According to Eskey, knowledge of form is linguistic in nature, and it includes recognition of graphophonics, lexical, syntactic/semantic and
rhetorical patterns of language. Knowledge of substance, on the other hand, encompasses cultural, pragmatic and subject-specific information and it provides the reader with certain expectations about the larger conceptual structure of the text. Whereas formal knowledge facilitates making accurate identification of forms from a minimum number of visual cues, subject-specific, cultural and pragmatic knowledge determines not only a personal reconstruction of the meaning of a text but also its depth and richness (Eskey, 1986). In Rumelhart’s (1994) ‘Toward an interactive model of reading’, the knowledge of form is classified into syntactic, semantic, orthographic and lexical knowledge. In an analysis of knowledge, Bernhardt (1991) identifies two types of knowledge: domain-specific knowledge and culture-specific knowledge (pp. 96–97), the combination of which coincides with Eskey’s knowledge of substance. Generally speaking, reading researchers often differ in their subdivisions of the reader’s prior knowledge. Their categorisations, however, demonstrate a tendency to make distinctions among linguistic, conceptual and sociocultural knowledge.

Like L1 (first language) reading comprehension, L2 (second or foreign language) reading comprehension starts with word recognition. Studies of eye movements in reading indicate that, as reported in Adams’ study (1990) and Bernhardt’s research (1991), readers do not guess or sample texts. Instead, they see almost every word in the text. According to Singer (1994), experienced readers differ remarkably from inexperienced readers in that they have attained automaticity in word recognition and usually focus their attention on higher-level processing, while inexperienced readers usually have to undertake fretful decoding, focusing their attention on the processing of the components of the individual words. Stanovich (1991) also notes:

‘[i]t is not that the good reader relies less on visual information, but that the visual analysis mechanisms of the good reader use less capacity. Good readers are efficient processors in every sense.’ (p. 21)

Being able to recognise L2 words, however, does not necessarily mean being able to interpret their meaning. L2 readers may well be able to recognise L2 words without knowing what they mean. They need to build up their vocabulary and learn to automate the link between word recognition and word meaning. As we all know, words usually do not come in isolation in natural texts, and their combination into larger units is governed by the syntax of the language (Lyons, 1981). Readers need syntactic knowledge to construct an interpretation of what they read. According to Adams (1980), ‘syntactic competence is an important dimension of linguistic competence in general’ (p. 18). Berman (1984) also notes, on the basis of her experiments with Hebrew-speaking college students learning English as a foreign language, that ‘efficient FL [i.e. fluent] readers must rely in part on syntactic devices to get at text meaning’ (p. 153).

Apart from word recognition and syntactic knowledge, L2 readers also need knowledge of textual structure. According to Meyer (1984), textual structure is realised in various linguistic devices such as superordinate references like topic sentences. It may also take the forms of ‘pointer words or signals’ (Roller, 1990), which indicate sequential, causal or concessive relations, etc. A considerable amount of research has been conducted about the impacts of textual organisation on L2 reading (Barnitz, 1985; Bernhardt, 1991; Carrell, 1984a; Cohen et al, 1979; Davies, 1984; Roller, 1990; Steffensen, 1987, 1988; Ulijn & Salager-Meyer, 1998; Urquhart,
Apart from knowledge of form, knowledge of content also influences L2 reading comprehension. According to Carrell (1983), ‘comprehending words, sentences, and entire texts involves more than just relying on one’s linguistic knowledge’ (p. 557), because ‘a text only provides directions for listeners or readers as to how they should retrieve or construct meaning from their own, previously acquired knowledge’ (Carrell, 1984b, p. 332).

The relationships between knowledge of content and knowledge of form in regard to their contributions to L2 reading comprehension have long been a research topic. According to Grabe (1991), most academically-oriented L2 learners are literate in one language and trying to become literate in another. Cummins (1976) and Clarke (1980) propose the notion of threshold level, as they find out that L2 readers will not be able to read as well in L2 as in their L1 until they have reached the threshold level of competence in L2. Ulijin and Salager-Meyer (1998) note in their studies that helping students of low L2 proficiency to improve their word-identification skill is a more important goal than helping them to develop problem-solving skills such as using context to figure out interpretations, intentions and conclusions. Hudson’s (1988) experimental data also show that L2 readers with advanced proficiency levels had less trouble processing visual information than those with lower proficiency level and that advanced L2 readers were able to bring more prior knowledge to bear on the reading process than beginning or intermediate L2 readers. Some researchers (Bensoussan, 1998; Carrell, 1988) argue that L2 readers’ inadequate L2 proficiency may cause them to over-rely on their prior conceptual and sociocultural knowledge to compensate for their insufficient L2 proficiency. Pritchard (1990) found that L2 readers appeared to use different reading strategies to read familiar and unfamiliar texts. Comparing proficient and less proficient ESL (English as a Second Language) readers in the United States, Fitzgerald (1995) concludes that:

‘[o]n the whole, more proficient ESL readers (a) made better use of vocabulary knowledge, (b) used a greater variety of metacognitive strategies and used selected strategies more frequently, (c) took more action to solve miscomprehension and checked solutions to problems more often, (d) used psycholinguistic strategies that were more meaning-oriented, (e) used more schema knowledge, and (f) made better and/or more inferences.’ (p. 180)

Research questions

The present research project stemmed from the need to find out what English reading materials are of appropriate readability for EFL students in China, as the EFL education system in China requires a rigid control of the readability of reading materials. Knowledge of the relationship between readers’ prior knowledge and their reading comprehension may help finding an efficient method to identify appropriate readability for students at different levels.

The study of the impacts of L2 readers’ prior knowledge on reading comprehension has usually been conducted through experiments about L2 readers’ reading
performance. Little has been reported in literature about L2 learners’ perceptions of their prior knowledge in terms of its contributions to reading comprehension. Although reading research has shown that linguistic, conceptual and sociocultural knowledge varies in their contributions to reading comprehension, it is little known how such variations are perceived by the readers themselves.

On the basis of Eskey’s (1986), Rumelhart’s (1994) and Bernhardt’s (1991) analyses of prior knowledge, the objectives for this research project are to find out 1) what prior knowledge has been perceived to be the most important for EFL reading comprehension, 2) whether EFL learners’ perceptions of the most important prior knowledge vary as their EFL proficiency varies, and 3) how they vary if such variations do occur.

**Method**

To address the three questions, a survey was conducted among 400 EFL students in a capital city in China by means of a questionnaire. The choice of using a questionnaire was mainly due to its high reliability, low cost, and ideal feasibility for the current situations in the participating educational institutions. To ensure the best possible validity, the questionnaire synthesised Eskey (1986), Rumelhart (1994) and Bernhardt’s (1991) analyses of prior knowledge and divided it into three categories. These constitute the underlying theory of the reader’s prior knowledge adopted for the investigation, namely the reader’s prior linguistic, conceptual and sociocultural knowledge is crucial to reading comprehension, though with varying effects. In order to accommodate responses that might involve factors different from the three categories, an extra category of ‘other factors’ was added.

It must be noted, however, that conceptual knowledge, in a strict sense, is usually acquired in a socio-cultural context and is thus always tinted socio culturally. It is difficult to define exactly when an instance stops being a member of conceptual knowledge and becomes a member of sociocultural knowledge and vice versa. The distinction made in the investigation between conceptual knowledge and sociocultural knowledge was meant to bring out a clearer picture of the prior knowledge involved in EFL reading comprehension.

**Subjects**

The survey involved four educational institutions, two at the secondary level and another two at the tertiary level. Following the usual ethical protocols, these educational institutions are referred to by codes rather than their original names: Middle School 1, Middle School 2, University 1 and University 2 (see Table 1 for details).

The 400 subjects were culturally and linguistically homogeneous in the sense that they were all Chinese, that they were all residents of the same geographical region and that they were all native speakers of Chinese. But they were also heterogeneous for three reasons: they were different in age, ranging from 13–23; they were studying different years at secondary and tertiary educational institutions, ranging from Junior Year 1 at middle school to Year 3 at university; their linguistic, conceptual and sociocultural knowledge differed substantially, from the lowest to the
approximately highest in terms of the EFL course design. The homogeneity of the subjects provided the necessary basis for the generalisation of the findings to the population from which the sample was drawn. The heterogeneity of the subjects, on the other hand, contributed to the representativeness of the sample.

In China, educational institutions are usually categorised into two groups: ordinary and key. Ordinary and key educational institutions set different entry scores and enjoy different levels of funding from the government. Middle School 1 is one of the many ordinary middle schools in the city, while Middle School 2 is one of the key middle schools. Similarly, University 1 is an ordinary university, while University 2 is a key university. The selection of students from both ordinary and key educational institutions, therefore, further ensured the representativeness of the sample.

The questionnaire

The questionnaire was in Chinese (i.e. the subjects’ L1) and it consisted of four multiple-choice questions, all accepting multiple responses. The subjects were asked to rank their multiple responses in order from most important to least important, and they were also invited to provide their own answers if they found the choices provided failed to describe their own situation.

To ensure comprehensibility, special care was taken to make sure that the questionnaire used only notions that the subjects were familiar with, and inquired only about what the subjects experienced in their EFL reading comprehension, presuming no prior special knowledge. The interviews of three experienced EFL middle-school teachers and another three university EFL teachers showed that the best way to survey the students’ perceptions of the roles of prior knowledge in EFL reading comprehension was to ask questions related to the difficulty they had encountered, and that all the six teachers believed that the cause of the difficulty in EFL students’ reading comprehension was causally related to their prior linguistic, conceptual and sociocultural knowledge. Consequently, the questions in the questionnaire were all targeted, directly or indirectly, at the cause(s) of the difficulty that EFL students might encounter in reading.

In order to minimise errors in representing the subjects’ perceptions of the cause(s) of the difficulty they had encountered in reading, the technique of triangulation was adopted. Four different questions were used to explore the same issue from four

<table>
<thead>
<tr>
<th>Institution ID</th>
<th>Year Level</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Middle School 1</td>
<td>Junior Year 1</td>
<td>76</td>
</tr>
<tr>
<td>Middle School 1</td>
<td>Senior Year 1</td>
<td>83</td>
</tr>
<tr>
<td>Middle School 2</td>
<td>Senior Year 1</td>
<td>97</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>256</td>
</tr>
<tr>
<td>University 1</td>
<td>Year 1</td>
<td>27</td>
</tr>
<tr>
<td>University 2</td>
<td>Year 1</td>
<td>41</td>
</tr>
<tr>
<td>University 1</td>
<td>Year 2</td>
<td>32</td>
</tr>
<tr>
<td>University 2</td>
<td>Year 3</td>
<td>44</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>144</td>
</tr>
</tbody>
</table>

Table 1. The breakdown of the subjects participating in the survey

© United Kingdom Reading Association 2002
different angles. Each of the questions was followed by a list of multiple choices designed to investigate the subjects’ conscious or subconscious perceptions of the roles that prior linguistic, conceptual and sociocultural knowledge might play in their reading comprehension.

It must be noted that, for many Western students, some of the terms used in the survey may seem rather technical, requiring some specialised knowledge for a full understanding, such as *idiom, syntax, formal structure* and *sociocultural background knowledge*. But these terms are not likely to cause problems with the participating subjects. The Chinese equivalents of these terms are introduced early in primary schools and they are frequently used and tested and have become part of the subjects’ working vocabulary. According to *The curriculum of Chinese for primary schools* (SEC, 1992a) and *The curriculum of Chinese for nine-year full-time compulsory education: Junior middle schools* (SEC, 1992b), the students in Year 2 in primary schools are expected to be able to identify homophones, polysemy and semi-homographs and the students in Year 1 in junior middle schools are expected to be able to undertake a structural analysis of sentences, identifying grammatical components such as subject, predicate, object, attribute, adverbial, complement, simple sentence, different types of complex sentence. Of all these terms used in the questionnaire, only one was later found to cause some problem among some middle-school students, that is, *idiom*, which does not have a counterpart in the Chinese language.

In the questionnaire, the first question was a general one. It was designed to identify the cause(s) that the subjects believed might give rise to the difficulty they encountered in reading comprehension. The subjects were given the following six choices.

- a) unknown EFL vocabulary
- b) too complicated syntax and/or formal structure
- c) unfamiliar content
- d) lack of the necessary sociocultural background knowledge
- e) unusual sentence length
- f) other factors such as: (Please specify)

Among the six choices, choices a, b and e concerned the reader’s linguistic knowledge. Choices b and e were similar, because choice e could serve as a more appropriate choice in case the subjects did not have sufficient comprehension of choice b. In addition, the subjects were also provided with an opportunity (i.e. choice f) to specify their own answer, if their perceived cause of difficulty was not identified in the list of choices.

The second question was intended to investigate how the subjects perceived factors other than vocabulary. It is not uncommon that an EFL reader cannot understand a text that contains no unknown words or expressions. It is also common that an EFL reader can recognise all the individual words used in a text, but is unable to produce a meaningful interpretation of the text. The second question was thus meant to identify the perceived major cause(s) for such a failure in reading comprehension. According to Grabe (1991) and Jiang (2000), the prior linguistic knowledge of L1 tends to encourage L2 learners to treat L2 words simplistically as ‘a second label for a well-understood concept’ (Grabe, 1991, p. 387), hence the problem of polysemy. Question 2, therefore, included a choice targeting specially the factor of polysemy,
although the question was meant to exclude the prior knowledge of vocabulary. The seven choices for question 2 are as follows.

a) unable to decide on the sense in which a word is used when the word may be used in several different senses
b) unable to understand the idioms in the text
c) unable to understand the subject content
d) lack of the necessary sociocultural background knowledge
e) unable to understand the syntax and/or the formal structure used in the text
f) unusual sentence length
g) other factors such as: (Please specify)

Of the seven choices, choices a, e and f concerned the reader’s linguistic knowledge. Choices c and d were aimed respectively at the reader’s conceptual and sociocultural knowledge. Choices e and f were redundant in the same way as they were in the previous question. Choice b was one that might be grouped together either with choice a or choice d. On the one hand, idioms have definite linguistic properties, for they have peculiar grammatical structure, functioning grammatically as a single word and cannot be understood from the individual meanings of their componential elements, for example, keep tabs on. The meaning of the idiom cannot be inferred from its componential lexical elements: keep, tab and on. On the other hand, however, idioms are also sociocultural in nature, for they are specialised expressions peculiar to a given social group. Since question 2 was meant to exclude the factor of EFL vocabulary, choice b was considered a choice concerning the reader’s prior sociocultural knowledge and grouped as such.

The third question was set to investigate the major factor(s) that might cause the EFL reader’s impaired performance in reading:

a) nervousness
b) difficult or boring content
c) shortage of time
d) lack of necessary sociocultural knowledge
e) other factors such as: (Please specify)

Unlike questions 1 and 2, the third question included only the choices that did not concern the reader’s linguistic knowledge. It included two factors that often caused impaired performance in reading, namely ‘anomaly of recall’ and ‘anomaly of attention’ (Reed, 1972). It is not uncommon to hear EFL students saying ‘When I went into the examination room I had it all pat. But as soon as I looked at the question papers my mind went a complete blank’. Such inability to recall what has been learned for optimal performance is an instance of ‘anomaly of recall’, which is usually referred to as ‘nervousness’ by Chinese students. ‘Anomaly of attention’ is another cause for impaired performance. When one is under the stress of time, one may become oversensitive in construing the present situation, continually vulnerable to anxiety about the time that is constantly ticking away. Chinese students usually refer to such an instance of ‘anomaly of attention’ as ‘shortage of time’ (see Reed, 1972, for detailed discussion of anomalies in performance).

The fourth and also the last question in the questionnaire concerned the opposite of what the first three questions addressed. It intended to find out the factors that the
subjects believed might make a text easy to understand. The following is the list of choices provided for the question:

a) absence of unknown EFL vocabulary  
b) simple EFL syntax  
c) interesting content  
d) familiar sociocultural background  
e) other factors such as: (Please specify)

Choices a–d in the list addressed respectively the reader’s prior linguistic (vocabulary and syntax), conceptual and sociocultural knowledge. In short, so far as the methodology triangulation is concerned, question 1 addressed the issue directly; question 2 targeted the same issue but tried to sift out the factor of the prior knowledge of vocabulary; question 3 went a step further to exclude linguistic knowledge; and question 4 intended to investigate the same issue but from the opposite direction.

The questionnaire was then pre-tested with the same six EFL teachers who were previously interviewed and was revised and shortened afterwards to a single page to facilitate the administration of the survey, and minimise the time required for the survey in class. For example, the choice of ‘too complicated syntax and/or formal structure’ for question 1 was a combination of two originally-separate choices. Similarly, the two original choices ‘difficult content’ and ‘boring content’ for question 3 were grouped into one, as they were both related to conceptual knowledge. The choice of ‘interesting content’ was condensed from three choices ‘easy content’, ‘interesting content’ and ‘familiar content’. However, no piloting was conducted for the questionnaire because of the administrative constraints. It was hoped that the large and representative sample that the survey involved might minimise the adverse impacts of the unanticipated errors in interpreting the questions posed in the questionnaire.

**Procedure**

The survey was conducted in class as a part of the teaching programme. The teacher who was engaged to conduct the survey in class gave a brief introduction to the purpose of the survey, advising the participating students that the present research project was about EFL reading comprehension, investigating the factors that Chinese EFL students believed would affect their reading comprehension of an English text. The students were also told of the estimated length of time required for completing the questionnaire, and that they could select more than one factor and were expected to rank their multiple-responses in order from most important to least important.

The subjects were also informed of the practical significance and benefits of the survey as well as their rights with regard to the project. They were encouraged to directly contact the principal researcher through the telephone number and postal address provided in the questionnaire. They were also asked to sign their declaration of consent should they agree to participate in the survey.

Common statistical analysis tools like SPSS and Minitab do not provide for an analysis of ranked multiple responses. A list of the frequencies of responses for each rank is difficult to interpret. A simple comparison of the observed frequencies of the selected ranks would fail to take into account the diminished probability for a
response to be selected after a decision has been made on the response to a higher rank. To solve the problem, *Multiplication Rule for Probability* is used to analyse the ranked multiple responses collected in the survey (see Appendix B for the formulae used in the analysis).

**Results**

*The ranked multiple responses to question 1*

The statistics of the ranked multiple responses to question 1 in Table 2a show that the most outstanding difference between middle-school students and tertiary EFL majors lies in the identity of the first factor. For middle-school students, it was choice b (i.e. EFL syntax and formal structure) but it was choice a (i.e. EFL vocabulary) for tertiary EFL majors. Furthermore, the ranking of choice b dropped from first position with middle-school students to fourth position with tertiary EFL majors. Such a drastic change in the ranking of choice b (i.e. EFL syntax and formal structure) may well indicate that EFL syntax and formal structures are more likely to provoke difficulty in reading comprehension at the initial stages of learning. As prior knowledge of English syntax and formal structures increases, students tend to find these aspects less difficult to understand. This does not mean, however, that the importance of the knowledge of English syntax and formal structures decreases with more advanced learners. What it may suggest instead is that students, having acquired the basic knowledge of English syntax and formal structure, may not be aware of their importance in reading comprehension, or that the importance of the basic knowledge of English syntax and formal structure may diminish for them.

Table 2a. The average expected values and expected relative frequencies of the ranked multiple responses to question 1 from EFL middle-school students and tertiary EFL majors

<table>
<thead>
<tr>
<th>Choice</th>
<th>Middle-school Students (n = 256)</th>
<th>Tertiary EFL Majors (n = 144)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E. value</td>
<td>R. freq.</td>
</tr>
<tr>
<td>a (Voc.)</td>
<td>5.025</td>
<td>0.308</td>
</tr>
<tr>
<td>b (Syn.)</td>
<td>6.731</td>
<td>0.422</td>
</tr>
<tr>
<td>c (Con.)</td>
<td>2.148</td>
<td>0.134</td>
</tr>
<tr>
<td>d (Soc.)</td>
<td>1.250</td>
<td>0.077</td>
</tr>
<tr>
<td>e (Sen.)</td>
<td>0.607</td>
<td>0.038</td>
</tr>
<tr>
<td>f (Oth.)</td>
<td>0.352</td>
<td>0.022</td>
</tr>
<tr>
<td>Total</td>
<td>16.112</td>
<td>1</td>
</tr>
</tbody>
</table>


© United Kingdom Reading Association 2002
If the responses in Table 2a are reorganised into linguistic, conceptual and sociocultural categories, then choices a, b and e may be subsumed into the category that concerns the reader’s prior linguistic knowledge. Choices c, d and f pertain respectively to the other three categories: conceptual knowledge, sociocultural knowledge and other factors. To offset the unequal numbers of the choices for the four categories, the averaged value is used for the category that includes more than one choice. Accordingly, the expected values and their relative frequencies of the four categories from middle-school students and tertiary EFL majors can then be tabulated in Table 2b.

It is obvious in Table 2b that the middle-school subjects considered linguistic knowledge as the most important source of difficulty in reading. But the tertiary EFL majors tended to regard sociocultural knowledge as the most important source.

It is also worth noting that middle-school students and tertiary EFL majors also tended to differ in their perceptions of the roles that conceptual and sociocultural knowledge performs in reading comprehension. Middle-school students seemed to consider conceptual knowledge more important than sociocultural knowledge (with a difference of 11.41%) while tertiary students seemed to consider sociocultural knowledge slightly more important (with a very small margin about 3.8%).

The ranked multiple responses to question 2

The seven choices for question 2 were all concerned with the cause(s) for the incomprehensibility of a text that contains no unknown EFL vocabulary. If the responses and their expected values are organised into the four categories they have been targeted at, then the statistics may be tabulated in Table 3.

It is obvious in Table 3 that, with the factor of vocabulary removed, the middle school students considered the reader’s prior conceptual knowledge to be the most
important for reading comprehension. But tertiary EFL majors tended to regard the reader’s prior sociocultural knowledge as more important.

**The ranked multiple responses to question 3**

Question 3 presumed an unusual situation when the readers felt that they could have done better than they did. The question was intended to find out what might cause impaired performance in EFL reading, if prior linguistic knowledge was not taken into consideration.

The feedback showed that the most likely cause for such impaired performance in reading was choice b (i.e. difficult or boring content). The next important factor was choice c (i.e. shortage of time). For the first time, middle-school students and tertiary EFL majors agreed on the ranking of the top two choices (see Table 4). Just as indicated in Table 4, the participating subjects, though differing considerably in their EFL reading abilities, tended to share the view that unusual content and environment were most likely to result in poor performance in EFL reading comprehension.

**The ranked multiple responses to question 4**

Question 4 was about the most important factor(s) that the subjects believed might make an EFL text easy to understand. Table 5a reveals an interesting difference in the ranking of choices b, c and d between middle school students and tertiary EFL majors. The lower the reader’s linguistic knowledge is, the higher the prior knowledge of EFL syntax and formal structures is ranked. In contrast, the better the reader’s linguistic knowledge is, the higher the reader’s prior conceptual and sociocultural knowledge is ranked. Choice b (i.e. the reader’s prior knowledge of EFL syntax and formal structures) was the first choice for middle-school students, but it was ranked as the lowest by most of the tertiary EFL majors, if disregarding the fifth choice.
concerning the factors not included in the list. Interestingly, the importance of choices c and d also increased considerably with tertiary EFL majors.

Again, if reorganised into the four categories, the responses to question 4 may be tabulated in Table 5b.

It is obvious in Table 5b that the middle-school students who participated in the survey believed that it was linguistic factors that were most likely to make a text easy to understand. But the tertiary EFL majors considered that it was the conceptual content that was most likely to make a text easy to understand.

Table 4. The average expected values and expected relative frequencies of the ranked multiple responses to question 3 from EFL middle-school students and tertiary EFL majors

<table>
<thead>
<tr>
<th>Choice</th>
<th>Middle-school students (n = 256)</th>
<th>Tertiary EFL Majors (n = 144)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E. value</td>
<td>R. freq.</td>
</tr>
<tr>
<td>a (Rec.)</td>
<td>3.638</td>
<td>0.182</td>
</tr>
<tr>
<td>b (Con.)</td>
<td>6.669</td>
<td>0.310</td>
</tr>
<tr>
<td>c (Att.)</td>
<td>6.365</td>
<td>0.293</td>
</tr>
<tr>
<td>d (Soc.)</td>
<td>3.519</td>
<td>0.167</td>
</tr>
<tr>
<td>e (Oth)</td>
<td>0.942</td>
<td>0.047</td>
</tr>
<tr>
<td>Total</td>
<td>21.132</td>
<td>1</td>
</tr>
</tbody>
</table>


concerning the factors not included in the list. Interestingly, the importance of choices c and d also increased considerably with tertiary EFL majors.

Again, if reorganised into the four categories, the responses to question 4 may be tabulated in Table 5b.

It is obvious in Table 5b that the middle-school students who participated in the survey believed that it was linguistic factors that were most likely to make a text easy to understand. But the tertiary EFL majors considered that it was the conceptual content that was most likely to make a text easy to understand.

Table 5a. The average expected values and expected relative frequencies of the ranked multiple responses to question 4 from EFL middle-school students and tertiary EFL majors

<table>
<thead>
<tr>
<th>Choice</th>
<th>Middle-school students (n = 256)</th>
<th>Tertiary EFL majors (n = 144)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E. value</td>
<td>R. freq.</td>
</tr>
<tr>
<td>a (Voc.)</td>
<td>6.059</td>
<td>0.274</td>
</tr>
<tr>
<td>b (Syn.)</td>
<td>7.415</td>
<td>0.346</td>
</tr>
<tr>
<td>c (Con.)</td>
<td>4.811</td>
<td>0.225</td>
</tr>
<tr>
<td>d (Soc.)</td>
<td>2.837</td>
<td>0.129</td>
</tr>
<tr>
<td>e (Oth)</td>
<td>0.556</td>
<td>0.026</td>
</tr>
<tr>
<td>Total</td>
<td>21.677</td>
<td>1</td>
</tr>
</tbody>
</table>


© United Kingdom Reading Association 2002
As Questions 1, 2 and 4 were all targeted at EFL readers’ perception of their linguistic, conceptual and sociocultural knowledge in relation to their contributions to reading comprehension, the responses they elicited may thus be synthesised and presented globally in Tables 6a and 6b.

Table 5b. The average expected values and expected relative frequencies of the ranked four categorical responses to question 4 from EFL middle-school students and tertiary EFL majors

<table>
<thead>
<tr>
<th>C. of c.</th>
<th>Middle-school students (n = 256)</th>
<th>Tertiary EFL majors (n = 144)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E. value</td>
<td>R. freq.</td>
</tr>
<tr>
<td>Ling. (a, b)</td>
<td>6.737</td>
<td>0.451</td>
</tr>
<tr>
<td>Con. (c)</td>
<td>4.811</td>
<td>0.322</td>
</tr>
<tr>
<td>Soc. (d)</td>
<td>2.837</td>
<td>0.190</td>
</tr>
<tr>
<td>Oth. (e)</td>
<td>0.556</td>
<td>0.037</td>
</tr>
<tr>
<td>Total</td>
<td>14.941</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: 1) E. value: Expected values 2) R. freq.: Relative frequencies 3) C. of c.: Category of choices. The letter(s) in the brackets are the identities of the choices 4) Ling.: Category concerning linguistic knowledge 5) Con.: Category concerning conceptual knowledge 6) Soc.: Category concerning sociocultural knowledge 7) Oth.: Category concerning other factors

A synthetic presentation of the results

As Questions 1, 2 and 4 were all targeted at EFL readers’ perception of their linguistic, conceptual and sociocultural knowledge in relation to their contributions to reading comprehension, the responses they elicited may thus be synthesised and presented globally in Tables 6a and 6b.

Table 6a. The average expected values and expected relative frequencies of the ranked four categorical responses to questions 1, 2 and 4 from EFL middle-school students and tertiary EFL majors

<table>
<thead>
<tr>
<th>C. of c.</th>
<th>Middle-school students (n = 256)</th>
<th>Tertiary EFL majors (n = 144)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E. value</td>
<td>R. freq.</td>
</tr>
<tr>
<td>Ling.</td>
<td>4.506</td>
<td>0.457</td>
</tr>
<tr>
<td>Con.</td>
<td>2.744</td>
<td>0.278</td>
</tr>
<tr>
<td>Soc.</td>
<td>2.282</td>
<td>0.231</td>
</tr>
<tr>
<td>Oth.</td>
<td>0.338</td>
<td>0.034</td>
</tr>
<tr>
<td>Total</td>
<td>9.871</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: 1) E. value: Expected values 2) R. freq.: Relative frequencies 3) C. of c.: Category of choices 4) Ling.: Category concerning linguistic knowledge 5) Con.: Category concerning conceptual knowledge 6) Soc.: Category concerning sociocultural knowledge 7) Oth.: Category concerning other factors

© United Kingdom Reading Association 2002
Table 6b. The breakdown of the average expected frequencies and expected relative frequencies of the four ranked categorical responses to questions 1, 2 and 4 from EFL middle-school students and tertiary EFL majors

<table>
<thead>
<tr>
<th>C.of c.</th>
<th>Middle-school students</th>
<th>Tertiary EFL majors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Junior Year 1 (n = 76)</td>
<td>Senior Year 1 (n = 180)</td>
</tr>
<tr>
<td>Ling.</td>
<td>4.075</td>
<td>0.454</td>
</tr>
<tr>
<td>Con.</td>
<td>2.677</td>
<td>0.298</td>
</tr>
<tr>
<td>Soc.</td>
<td>1.870</td>
<td>0.208</td>
</tr>
<tr>
<td>Oth.</td>
<td>0.357</td>
<td>0.040</td>
</tr>
<tr>
<td>Total</td>
<td>8.979</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: 
1) E. value: Expected values 
2) R. freq.: Relative frequencies 
3) C. of c.: Category of choices 
4) Ling.: Category concerning linguistic knowledge 
5) Con.: Category concerning conceptual knowledge 
6) Soc.: Category concerning sociocultural knowledge 
7) Oth.: Category concerning other factors

United Kingdom Reading Association 2002
Discussion

The survey conducted among 400 Chinese middle-school students and tertiary EFL majors gave a fairly clear picture of their perceptions of prior linguistic, conceptual and sociocultural knowledge in terms of respective contributions to reading comprehension. The data indicated two different tendencies for the two participating groups of subjects. The middle-school students tended to give the top ranking to the choices concerning either the reader’s linguistic knowledge or the linguistic characteristics of the text; the tertiary EFL majors, however, were likely to give the top priority to the choices that concern either the reader’s prior conceptual and sociocultural knowledge or the conceptual and sociocultural background knowledge presumed by the text.

The different ranking of the first choice shown in Table 6a between middle-school students and tertiary EFL majors can be accounted for if taking into consideration the difference in the prior linguistic knowledge between the two groups. Using expected vocabulary range as an index, for example, middle-school graduates who have completed a six-year EFL language course are expected to acquire 3000 words, while graduating tertiary EFL majors are expected to have a vocabulary of 9000–12000 words (S.E.C., 1992b, 1993; Design Group, 1990). It should not be too difficult to infer from the notable difference in vocabulary range that the two groups of students differ considerably in their linguistic knowledge.

Taking a closer look at the relations among the breakdowns in the previous tables, especially in Table 6b, we can see that the perceived importance of vocabulary, syntax and formal structures decreases as the subjects’ year-level in the EFL course increases. But the perceived importance of conceptual and sociocultural knowledge increases along with the increase of the subjects’ year-level in the EFL language course. In other words, so far as EFL reading comprehension is concerned, it seems that the more EFL vocabulary, syntax and formal structures are learned, the less importance is attached to them. As this occurs, greater importance is attached to conceptual and sociocultural knowledge.

The inverse relation between the perceived importance of linguistic knowledge and that of conceptual and sociocultural knowledge in terms of their respective contributions to reading comprehension is clearly revealed in Table 6b. The ranking of the category of linguistic factors drops from first position for middle-school students to second position with tertiary EFL Year-1 majors and further down to third position with tertiary EFL Year-2 and Year-3 majors.

The marked difference in these Chinese EFL learners’ ranking of the top category of factors indicates that, so far as L2 reading comprehension is concerned, linguistic knowledge of L2 is of crucial importance for beginners as well as for those whose L2 competence has not reached the threshold level, but it is not necessarily so with more advanced L2 learners. The point where sociocultural or conceptual knowledge replaces linguistic knowledge as the most important and decisive factor seems to fall somewhere between the end of secondary education and the initial period of tertiary education, when EFL learners have acquired a vocabulary of 3000 words and the basic syntax and formal structures of English (S.E.C. 1992b, 1993). This finding agrees with what Laufer (1992) notes about L2 threshold level.

‘As for the nature of this language competence ceiling, it is believed to be largely lexical (Ulijn, 1984; Ostyn et al., 1987; Laufer and Sim, 1985a and 1985b) and
the number of words constituting the threshold has been suggested to be 3,000 word families, or 5,000 lexical items.’ (pp. 96–97)

In summary, as an answer to the three research questions, the investigation indicates that middle-school students perceive the reader’s prior linguistic knowledge as the most important factor for EFL reading comprehension, while tertiary EFL majors consider the reader’s prior sociocultural knowledge as the most important. The survey also shows that, as readers’ prior linguistic knowledge increases, they tend to attach less importance to it, especially the knowledge of syntax and formal structures. At the same time, conceptual and sociocultural knowledge assumes greater importance. The perceived importance of prior linguistic knowledge seems to start diminishing around the end of secondary education when the students have acquired a vocabulary of 3000 words and the basic syntax and formal structures of English. The final replacement of linguistic knowledge by sociocultural knowledge as the top factor affecting reading comprehension seems to take place about one year after beginning tertiary EFL education. This may well be considered as an indication of the EFL threshold level for EFL students in China.

Two observations about the survey need to be noted. First, the survey was about the prior knowledge that EFL learners believe is involved in reading comprehension. It is possible, no matter how slightly it might be, that students’ perceptions deviate from their actual performance in EFL reading and thus fail to provide a true account of their reading comprehension. Second, the questionnaire may be said to be biased against factors other than linguistic, conceptual and sociocultural ones, as few other factors have been included in the survey. Although it could be argued that the subjects were invited to specify the factors that they thought to be influential but were not included in the questionnaire, few actually took the trouble to supply them.

Finally, the conclusions inferred from the survey seem to provide additional support to the L2 Linguistic Threshold Hypothesis: that L2 reading comprehension is both a matter of language and a matter of reading. With beginning L2 readers, reading comprehension is more a matter of language than a matter of reading. The findings of the survey also cast sidelight on the issue of the relations between the knowledge of L1 and L2. It seems that the issue in L2 reading comprehension education is not whether L2 language teachers should tap into L2 learners’ L1 reading ability and L1-related knowledge, but how to tap into it to secure a maximisation of the positive L1 transfer.

Appendix A. A questionnaire on English reading comprehension

(Translation from Chinese)

1) The major factors that cause difficulty in reading comprehension are: (You may choose more than one factor. Please rank the factors you choose in order from most important to least important.)

A) Unknown EFL vocabulary.
B) Too complicated syntax and/or formal structure.
C) Unfamiliar content.
D) Lack of the necessary sociocultural background knowledge.
E) Unusual sentence length.
F) Other factors such as: (Please specify.)

2) The English text, which contains no new words, might not be comprehensible, mainly because of the following factor(s): (You may choose more than one factor. Please rank the factors you choose in order from most important to least important.)

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
A) Unable to decide on the sense in which a word is used when the word may be used in several different senses.
B) Unable to understand the idioms in the text.
C) Unable to understand the subject content.
D) Lack of the necessary sociocultural background knowledge.
E) Unable to understand the syntax and/or the formal structure used in the text.
F) Unusual sentence length.
G) Other factors such as: (Please specify.)

3) The major factor(s) that could cause poor performance in reading comprehension are: (You may choose more than one factor. Please rank the factors you choose in order from most important to least important.)

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
A) Nervousness.
B) Difficult or boring content.
C) Shortage of time.
D) Lack of necessary sociocultural knowledge.
E) Other factors such as: (Please specify.)

4) Suppose you read an English text or sentence for the first time. If you find it easy to understand, it is mainly because of the following factor(s): (You may choose more than one factor. Please rank the factors you choose in order from most important to least important.)

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
A) Absence of unknown EFL vocabulary.
B) Simple EFL syntax.
C) Interesting content.
D) Familiar sociocultural background.
E) Other factors such as: (Please specify.)

Appendix B: Formulae for the analysis of ranked multiple responses

The following are the formulae of Multiplication Rule for Probability (1a–b) and Expected Value of a Discrete Random Variable (1c) used in the analysis of ranked multiple responses is as follows
For any events \( X \) and \( Y \),
\[
P(X \text{ and } Y) = P(X|Y) \cdot P(Y) \quad (1a)
\]
and
\[
P(X \text{ and } Y) = P(Y|X) \cdot P(X) \quad (1b)
\]

\[
E(X) = \sum_{i=1}^{n} x_i P(X = x_i) = \sum_{i=1}^{n} x_i P(x_i) \quad (1c)
\]

In Formulae 1a–b, \( X \) and \( Y \) are events respectively representing the selection of a particular choice provided in the list. \( P(X) \) and \( P(Y) \) are probabilities for Event \( X \) and Event \( Y \), respectively. \( P(X|Y) \) is a conditional probability of Event \( X \) when Event \( Y \) has occurred. \( P(X \text{ and } Y) \) is a joint probability for the joint event (\( X \) and \( Y \)) to occur. In Formula 1c, \( E(X) \) is the expected value for the random variable \( X \), which, in the survey, is the expected frequency for a given rank; \( x_i \) is the observed frequency of a given choice being selected for the rank; \( P(X = x_i) \) is the probability of the value \( x_i \).

References


Address for correspondence: Dr Zheng Lin, School of Education (Mt Helen Campus), University of Ballarat, VIC 3352, Australia.