A Corpus-based Study of Collocation Production of Chinese EFL Learners

Bowen Gao[1,*], Lin Zhu[1] & Ziyan Luo[1]

[1]School of Foreign Languages, Xi'an Aeronautical University, China

*Corresponding author: Teaching assistant of of School of Foreign Languages, Xi'an Aeronautical University, China. E-mail: 201903002@xaau.edu.cn

Abstract

It is of great difficult for EFL (English as a Foreign Language) learners to acquire collocation. This research investigated V + N collocation production of Chinese middle school university EFL learners based on TECCL (Ten-thousand English Composition of Chinese Learners) corpus, aiming at discovering the difficulties in collocation usage. In the study, collocations in the writing samples were identified manually and evaluated by dictionaries, corpus and native speakers as right or wrong. Collocation accuracy was figured out and errors were examined. Therefore, difficulties in collocation usage were discovered. The data were analyzed both quantitatively and qualitatively to yield the following three findings: first, middle school students and university students respectively produced 79% and 83% correct collocations among their collocation production. Second, collocation accuracy did not improve significantly as overall L2 (second language) proficiency increases. Third, wrong choice of verb and noun was the largest problem for Chinese EFL learners. The results also suggested that L1 (first language) influence and deficiency in L2 knowledge affected collocation usage. Based on these findings, pedagogical implications were discussed.

Keywords: corpus-based, collocation acquisition, Chinese EFL learners

1. Introduction

Lexicon and grammar are both significant building blocks in language. However, most EFL (English as a Foreign Language) teachers pay more attention to grammar than vocabulary, and ignore the role of the lexicon. The attention on the lexical instruction derived from Lewis's (1993) groundbreaking view that 'language consists of grammaticalised lexis, not lexicalised grammar' (Lewis, 1993, vi). From then on, vocabulary has become more and more important.

In the field of lexicon, researchers pointed out that fixed and semi-fixed phrases, such as idiom and collocation, should be granted as important roles in language development (Bahns, 1993; Sinclair, 1991). This view has been supported by scholars in the field of second language acquisition (Bahns and Eldaw, 1993; Cowie, 1998; Granger, 1998; Howarth, 1998), and researchers in pedagogical perspective (Ellis, 1996, 2001; McCarthy, 2004; Nation, 2001). They also supported that EFL/ESL (English as a Second language) learners at each level should be aware of the significance of collocation.

Therefore, it is of great interest to find out how students use collocation. Previous studies have found that the comprehension of collocation was of no difficulties for students (Lewis, 2000; Nesselhauf, 2003), however, the problem lied in the production of collocation (Nesselhauf, 2003). In studying the production of collocation, most researchers focused on elicited tasks such as filling in the blank task or translation task, but these techniques could not reflect the collocation in practice. In addition, the difficulties in the usage of collocation have not been investigated in detail.

The aim of this research was to figure out to what certain extent would Chinese EFL learners produce appropriate collocation, and to clarify the problems in the usage of collocation. Based on TECCL (Ten-thousand English Composition of Chinese Learners) corpus, the research focused on the V-N collocations (such as 'fail the exam' or 'acquire the knowledge') in Chinese EFL learners' written production.

2. Literature Review

2.1 Theoretical Background

The initial attention on collocation derived from EFL/ESL teachers' suggestion that collocation instruction should be included in the class (McCarthy, 1984; Nation, 1990; Nattinger, 1980). The lexical approach which was granted as an important role in word and word combinations (Lewis, 1993) emerged as required. Lexical approach emphasized that the essence of language teaching lay in language chunks and multi-word units (Richards and Rogers, 2014; Schmitt, 2008). The goal of teaching was to develop students' ability to process and produce language chunks (Boers and Lindstromberg, 2009; Lewis, 1993, 1997). Schmitt (2008, cited in Richard and Richards and Rogers, 2014) further commented lexical approach:

'...teaching language based on the idea that language is made up of lexical units rather than grammatical structures...the units are words and chunks formed by collocations and fixed phrases.'

With the development of lexical approach, the significance of language chunks became undisputed (Nation and Newton, 1997).

However, researchers (Lewis, 2000; Schmitt, 1999; Schmitt, 2000) suggested that collocations created great difficulties for Second language learners (L2) learners from the pedagogical perspective. Some studies provided evidence that even advanced L2 learners had problems in using collocations (Fan, 2009; Nesselhauf, 2003; Nesselhauf, 2005; Yamashita and Jiang, 2010). Therefore, it is of great necessity to investigate EFL/ESL learners' production of collocation.

2.2 Definition and Classification of Collocation

In this research, the approach employed to identify collocation was *phraseological approach* which took grammatical structure and pragmatic use of collocation into consideration (Cowie, 1994). Thus, it would be able to analyze the complex relationship among collocation elements and the context of collocation. Scholars who supported this approach (Cowie, 1994; Howarth, 1998; Nesselhauf, 2005) developed two criteria to classify the collocation: transparency and combinability. Transparency referred to whether the combinations could be translated literally, and combinability referred to 'whether and to what degree pragmatic substitution of the elements of the combination are restricted' (Cowie, 1981, cited in Peng, 2016: 12).

Based on these criteria, collocation was distinguished from 'free combinations', which was defined as word combinations whose elements could be substituted arbitrarily without changing their meaning, such as 'require something'; and from 'idiom' which referred to expressions whose element could never or hardly be changed or replaced, such as 'rain cats and dogs' (Gao and Zhang, 2005; Howarth, 1998; Nesselhauf, 2005; Peng, 2016).

Of the whole collocations, lexical and grammatical collocations were also distinguished by scholars (Nesselhauf, 2003; Peng, 2016). Lexical collocations only consist of lexical items such as 'fail the exam' or 'strong tea', while grammatical collocation referred to combinations with grammatical elements (prepositions or determiners), such as 'angry at' or 'interest in' (Nessehaulf, 2004).

In terms of the investigation of V-N combination, scholars further developed methods to distinguish collocation from other types of combinations. For example, Gao and Zhang (2005: 106) developed a criterion to define V-N collocation based on whether the combination or its variance could be found in the dictionary 'under the entry of its verb and under the entry of its noun' (Gao and Zhang, 2005: 108). An example provided was 'acquire knowledge'. Look 'acquired' and 'knowledge' up in the dictionary *NCDCEU* (*The New Century Dictionary of Current English Usage)*, 'Acquire a working knowledge of English' was found in the explanation of word 'acquire', and 'We acquire knowledge step by step' in 'knowledge'. The combination occurred under the use of the verb and noun in the dictionary, thus its elements were seen to be mutually expected. Therefore, it was classified as a 'restricted collocation', which meant a strong association among the elements of collocation. If no indications could be found under the use of both the noun and verb in the dictionary, it was classified as a free combination, such as the combination 'go back home'.

Many linguists supported that general approach to measure the 'collocational restriction' was to use dictionary or corpus as basis, combined with native speaker judgements (Nesselhauf, 2004; Zhang and Gao, 2005; Peng, 2016). However, even under the same standard, definition and collocation from different researchers may not always be the same (Nesselhauf, 2003).

2.3 Review of Empirical Studies on L2 Collocation

2.3.1 Studies on the Correlation between Collocation Knowledge and Language Proficiency

The investigation on collocation acquisition had long been identified (Levenston, 1979; Lewis, 1993). The empirical research started from the 1990s.

First, the relationship between the acquisition of collocation and the language proficiency was studied by several researchers with correlational methodology (Al-Zahrani, 1998; Hsu, 2008; Hsu and Chiu, 2008; Sung, 2003; Yazdandoost et al, 2014; Zhang, 1993). Yazdandoost et al (2014) explored the relationship between the knowledge of collocation and reading, writing, listening and speaking proficiency of Iranian EFL students. A multiple choice collocation test was used to elicit students' collocation knowledge and an IELTS (International English Language Testing System) test was administered in order to identify students' language proficiency. The data indicated a significant correlation between collocation knowledge and the four language skills.

The result was consistent with other studies associating the collocation and individual language skills. In Hsu's (2007) study, an online writing task was administered to measure students' writing fluency and meanwhile figure out the frequency and the variety of lexical collocation. Results of Pearson correlation showed that a strong correlation existed between the collocation frequency (r=0.544, p=.00) and writing scores, as well as the collocation variety and writing scores (r=0.809, p=0.00).

Chiu and Hsu (2003) had explored the correlation between the knowledge, usage of collocation and speaking proficiency. Learners' knowledge of lexical collocation was evaluated by a fill-in-the-blank collocation test, and the usage of collocation was measured by two spoken tasks. The findings suggested that there was a statistically significant correlation between the knowledge of lexical collocation and speaking proficiency (r = 0.561 at p<0.01 level). However, no strong relationship was found between the usage of lexical collocation and speaking proficiency, nor between the knowledge and usage of lexical collocation.

In sum, all the reviewed studies had described the correlations between knowledge of collocation and the language proficiency, and their results showed similar positive relations.

2.3.2 Studies on the Effect of Explicit Collocation Instruction

The effect of collocation instructions in classroom had also been investigated by quantities of researchers (Harley, 1994; Hsu, 2002; Lien, 2003; Schmidt, 1993; Tseng, 2002).

Mounya (2010) conducted an experiment to investigate the effect of teaching collocations in the development of foreign language writing proficiency. Participants were randomly assigned to two groups. Both groups received the same writing lessons, however, only the experimental group was taught collocation. A test after these treatment was administered to measure the writing scores. Results revealed that writing scores of the experimental group were higher than those of the control group, and a strong correlation existed between the students' writing proficiency and collocations usage (r=0.84). This indicated that teaching collocation was an effective way to improve writing proficiency.

Mounya's (2010) findings were in line with Lien's (2003) results. Lien (2003) conducted a four-week's experiment to examine the effect of direct collocation instruction on Taiwanese college English majors' reading comprehension. Three groups were consisted of students in three academic levels: sophomore, junior and senior. Three types of instructions were single

item vocabulary instruction, collocation instruction, and no instruction. The results indicated that collocation instruction caused more positive effect on the reading skill.

Looking at another specific language skill, Hsu (2005) explored the correlation between collocation instruction and the listening comprehension of Taiwanese university students. Same with Hsu's (2003) results, those who received collocation instruction achieved the highest scores of the listening test compared with students who received vocabulary instruction and no instruction.

Other studies concerning the effect of collocation instruction on speaking fluency (Movahediyan et al, 2013), vocabulary development (Lin, 2002; Tseng, 2002) and overall language proficiency (Hsu, 2002) has been investigated. The results similarly showed that direct instruction positively correlated with the language acquisition and could enhance the four language skills. However, the studies reviewed up only focus on the collocation knowledge and collocation performance in tasks till now.

2.3.3 Studies on Learners' Collocation Production

Studies on collocation production of L2 learners were generally concerned with the following three aims: to investigate how target collocations were used by EFL/ESL learners in written or spoken tasks (Chen and Baker, 2010; Durrant and Schmitt, 2009; Fan, 2009; Jiang and Yamashita, 2010; Tsai, 2015); to identify how L2 learners committed collocation errors (Nesselhauf, 2003; Nesselhauf, 2004; Yamashita and Jiang, 2010) and to figure out the factors which influence collocation usage (Nesselhauf, 2003; Yamashita and Jiang, 2010).

First, studies focusing on learners' collocation usage were generally accomplished by comparing the collocation competence between native and non-native speakers with comparable database. Overall results were that both EFL and ESL learners all 'underuse, overuse, or misuse collocations' in actual application (Henriksen, 2013, cited in Peng, 2016, p.22).

Tsai (2015) compared Taiwanese EFL learners' collocation usage in their written discourse with that of English native speakers using British National Corpus (BNC). The results showed striking differences between L2 learners and native speakers in number and type of collocations they used. L2 Learners produced 639 collocation types. By contrast, native speakers produced 1628. However, when comparing the collocation density, 19.35 collocation per 1000 words were found by native speakers, while 31.27 collocation per 1000 words by L2 learners (Tsai, 2015). This suggested that L2 learners tended to rely excessively on a few common collations, but rarely use low-frequency collocations.

Schmitt and Durrant (2009) examined the collocations found in native and non-native written texts, with International Corpus of Learner English (ICLE) and Louvain Corpus of Native English Essays (LOCNESS). The results also indicated that compared with native speakers, non-native speakers extensively used common and 'safe' collocations in high frequency. However, they rarely used 'less common ones and strong associated items', such as 'densely populated, bated breath and preconceived notions' (Durrant and Schmitt, 2009, p.175).

The results were in line with the study focusing on oral tasks. Serrano et al (2015) studied the formulaic sequences (language chunks) produced by three levels of Spanish EFL learners and native speakers. The findings also showed that significant differences still existed between advanced L2 learners and native speakers in the types of collocations (p<0.001).

Second, in terms of collocation errors, Nesselhauf (2003, 2004) studied near 2,000 V-N

collocations elicited from advanced German EFL learners' compositions. Nesselhauf (2003) evaluated the acceptability of collocations by dictionaries, corpus and native speakers. Collocations were judged as acceptable if they were found in dictionaries or the corpus, otherwise they were judged by the native speakers as right or wrong. The results indicated that approximately one-third of the collocation were considered questionable or problematic, and most errors were found in restricted collocations.

Gao and Zhang (2005) explored five levels of Chinese EFL learners' collocation production in compositions. They found that compared with lower level students, high level learners still made a lot of errors. They explained that the more advanced a collocations was, the more difficulties it were accompanied by. And thus, the more advanced collocation they used, the more errors occurred accordingly.

Of all types of collocations across languages, V-N lexical collocation was found the most problematic for L2 learners (Nesselhauf, 2005; Peng, 2016; Serrano et al, 2015), such as 'overcome difficulty'. This was due to their 'arbitrariness and unpredictability' (Gitsaki, 1996, cited in Peng, 2016: 19).

Third, with regarding the developmental process in L2 learners' collocation acquisition, studies showed that L2 learners' collocation knowledge expanded with the development of general English proficiency (Bahns and Eldaw, 1993; Ellis, 1993). Zhang (1993) investigated the acquisition development of collocation. The results indicated that students with lower level of English proficiency produced more grammatical collocations than lexical ones. When achieving advanced level, learners tended to use more grammatical ones. This was because that learners would be able to break down the grammatical collocations into parts and create new ones by themselves. However, he suggested more empirical evidence to support this finding.

Fourth, scholars (Fan, 2009; Shitu, 2015; Yamashita and Jiang, 2010) also clarified factors that affected learners' knowledge of collocation. Shitu (2015) studied advanced ESL learners' lexical collocation errors in essay writing and found that errors were significantly caused by learners' L1 transfer. Evidence that inadequate collocation knowledge and various instruction conditions might lead to collocation errors had been provided (Fan, 2009; Nesselhauf, 2003; Hsu and Chiu, 2008).

To sum up, research on collocation production of L2 learners described the differences in collocation use between native speakers and L2 learners, and confirmed that even advanced learners had difficulties in collocation use.

2.3.4 Summary of Empirical Research on L2 Collocation

In terms of the data used for analyzing learners' collocation usage, two types of research were categorized according to their methodologies employed: one with elicited data and the other with corpus. For the first type, various data could be elicited for different research goals. Collocation elicitation techniques ranged from translation task, cloze test, multiple choices to judgement tasks. However, research with these methodologies had not been mentioned in the current study. The second type of experiment used large scale learner corpora for analysis, such as LINDSEI and ICLE.

One of the downsides of elicitation tools was that learners produced single collocations without context, which led to the findings 'failed to reflect the actual performance of learners' in collocation usage (Fan, 2009: 112). Corpus analysis in this research avoided this defect, because the compositions in TECCL corpus were produced in free use in. The current research aimed at figuring out to what extent Chinese EFL learners use appropriate collocations and

further identifying the difficulties of natural collocation use. The corpus analysis fulfilled this purpose better.

Based on the literature reviewed up till now, most studies compared the collocation usage of native and non-native speakers. However, few concerned the EFL or ESL learners with different levels of proficiency. With regards to language learning context, only few focused on Chinese learners' collocation performance. Moreover, most of the studies reviewed concerned more on the context that learners would perform, such as collocation tokens and types, but these were rather descriptive. Only few concerned on how they performed, such as how they made errors (Nesselhasuf, 2003) and how the collocation acquired (Zhang, 1993). In this regard, further research on how L2 learners used collocation is worth carrying out. The present study will examine what and how collocation used by Chinese EFL learners: to what extent will Chinese EFL learners produce appropriate V-N collocation in free writing and how they make errors.

3. Research Design and Methodology

3.1 Research Questions

The research questions of the current research were:

- To what extent do Chinese EFL learners use appropriate verb-noun collocations?

- Does collocation accuracy improve as overall L2 proficiency increases?

- If collocation accuracy improves as L2 proficiency increases, how does the accuracy improve?

3.2 Methodologies

3.2.1 Data Source

Data source and data selection

The corpus data employed was TECCL (Ten-thousand English Compositions of Chinese Learners) corpus, which contains approximately 1000 written samples of Chinese EFL learners in different learning stages: from elementary school to post graduate. TECCL was used for a couple of reasons. First, the compositions were produced between 2011 to 2015, which reflects the up to date language acquisition phenomenon of Chinese learners. Second, the writers of compositions came from 32 provinces, and the geographically spread is the widest among Chinese EFL learner corpus. Third, samples which cover1000 essay topics came from tasks in and after class, and even collaborative work, which is different from most of the existing corpus that comprises texts from the examinations such as the CET-4/6 or TEM4/8.

The 22 compositions selected came from 11 middle school students and 11 university students. The samples were argumentative or narrative compositions, including the topic such as 'Reading selectively or extensively', 'Social network, bad or good?' and 'Should we help strangers?'. For the sake of random selection, criteria employed were: first, 22 compositions were selected from 20 provinces; second, a composition whose topic has been covered by a selected sample would be replaced by another one. The compositions written by middle school students were about 200 words, and that of university students were about 500. In order to make the samples more comparable concerning the length of the essay, university students' compositions were reduced to the first 200 words. This was justified because collocation

patterns are likely to be the same in the first and second halves of an essay.

3.2.2 Collocation Classification System

First, all verb-noun combinations were identified manually, and considered as being the active voice. For example, in 'the air has been polluted seriously by human beings', the V + N collocation 'pollute the air' would be extracted. If several combination were different instantiations of the same one in one essay, they were counted as the same one. For example, 'fails the exam' and 'failed the exam' are the instantiations of 'fail the exam', and therefore they were considered as one collocation.

The second step is to classify combinations according to the degree of restriction. Two dictionaries used were: the *Oxford Advanced Learners Dictionary (OALD 2000)* and the *Longman Dictionary of Contemporary English (LDCE 2005)*. These two dictionaries were found to contain detailed information of word combination (Nessehaulf, 2003). Based on definition of 'restriction' which mentioned before (Gao and Zhang, 2005; Nesselhauf, 2003), the procedures were as follows.

If one combination could not be found in two dictionaries, it was classified as free collocation. For example, 'require help'. The information in LDCE were:

Require: require something: to need something.

Help: things you do to make it easier or possible for someone to do something.

If one combination or its variation could be found in one dictionary under the entry of its verb and under the entry of its noun, the verb and the noun were considered mutually expected. Therefore, it was classified as collocation with high degree of restriction (CL1), such as 'achieve goal'. The following information in LDCE were found:

Achieve: She eventually achieved her goal of becoming a professor.

Goal: something you hope to achieve in the future.

If one combination occurred once in the use of either its noun or its verb, but it could be found in another dictionaries, it was also classified as a collocation with high degree of restriction(CL1), such as 'obtain permission':

OALD: obtain: to get something by making effort: to obtain permission

LDCE: obtain: You will need to obtain permission from the principal.

If only one occurrence could be found in two dictionaries, the combination was classified as collocation with medium degree of restriction (CL2), such as 'exert pressure':

LDEC: Environmental groups are exerting pressure on the government to tighten pollution laws.

If a combination could not be understood by its literal meaning at the time, substitution of any element was impossible, it was classified as an idiom. For example 'rain cats and dogs'

LDCE: to rain heavily.

Therefore, all V + N combinations were classified into four types, which were shown in Table 1.

Table 1. Combination Types and Criteria

Types of combination		Criterion	Example
Free combination (F)		No possible occurrence in dictionaries	Require help
Collocation	CL 1	Occur twice in two dictionaries	Obtain permission
	CL 2	Occur once in two dictionaries	Exert pressure
Idiom (I)		A combination cannot be understand by its literal meaning	Rain cats and dogs

3.2.3 Collocation Evaluation System

The third step was to evaluate the appropriateness of the combinations. Non-lexical elements such as modifiers, prepositions, determiners and complements were also taken into account, such as pay close attention 'to' and strike a balance 'between'. According to Nessehaulf (2003), five scales of acceptability were assumed: *clearly acceptable, largely acceptable, unclear; largely unacceptable and clearly unacceptable*. However, two problems arose: First, no clear distinction was made between wrong and largely wrong ones. Second, collocations that were judged as 'not sure' and 'OK, but not the best way to say it' were classified into one category--unclear. Thus, this type contained self-created collocations and existing collocations which were misused, which led to the unclear boundaries between each type.

Therefore, simpler and clearer procedures were employed. Collocations were correct if they could be found in any two of the following three dictionaries: *LDCE (2005), OALD (2000), The BBI Dictionary of English Word Combinations (1995)* or in three different texts in the BNC (British National Corpus). The collocation could not be judged as correct were evaluated by four native speakers. The recruitment of native speakers was based on a questionnaire (see appendix A).

Each native speaker were given ten compositions, they judged the combinations as 'correct'; 'wrong'; 'OK, they are acceptable, but not the best way to say it'; and '?' for those cannot be understood. Native speakers were also asked to give their own correction (detailed instructions for native speakers see appendix B).

If two informants judged 'correct', the collocation was assumed to be clearly acceptable. If both informants judged 'OK' or one 'correct' one 'OK', the collocation was classified as 'largely acceptable', otherwise, as unacceptable. Two native speakers judged for one case. Table 2 clearly shows the degree of acceptability and the criteria for each type.

Table 2. Degree of Acceptability and Criteria

Degree of acceptability	Criteria
Clearly acceptable	Found in any two dictionaries, in three different texts in the BNC, or judged as correct by two native speakers.
Largely acceptable	Both native speakers judged as 'OK' or one 'correct' one 'OK'
Unacceptable	Both native speakers judged as 'wrong'

This method has been tested to be effective and reliable: although native speakers' corrections for the wrong combinations were different, they agreed on around 96 percent correct ones. It should be noted that 'mistakes' and 'errors' were not differentiated in this study, and neither were 'correct' and 'acceptable', and 'wrong' and 'unacceptable'.

4. Results

4.1 Overall Findings

As noted earlier, verb-noun combinations were classified according to the degree of restriction and degree of acceptability. Altogether, 312 verb-noun combinations were identified from the samples, among which 207 were classified as free combinations, 105 collocations and no idioms. Table 3 presented the overall distribution of combinations on the scale of acceptability and restriction. Table 4 showed the distribution of combinations of middle school and university students.

Table 3. Overall Distribution (Degree of Restriction and Acceptability) of Combinations

| Types of collocation	Degree of acceptability			Total
	Clearly acceptable	Largely acceptable	Unacceptable	
Free combination	140	23	44	207
Collocation	73	10	22	105
Idiom	0	0	0	0
Total	213	33	66	312

Table 4. Distribution (Degree of Restriction) of Combinations in Writing Samples

Combination types	Middle school students	University students
free combinations	89	118
collocations	47	58
idiom	0	0

Figure 1 and 2 show the ratios of three types of combinations used by middle school students and university students respectively.
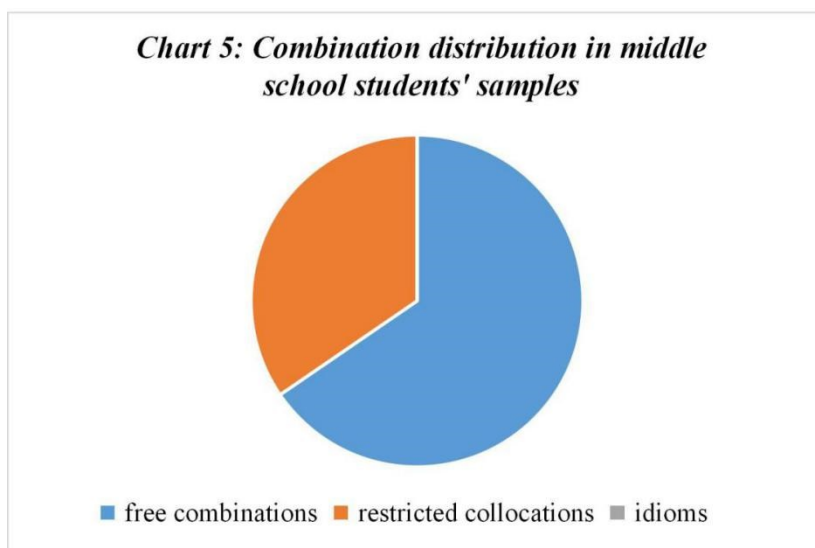
Figure 1. Combination Distribution in Middle School Students' Samples
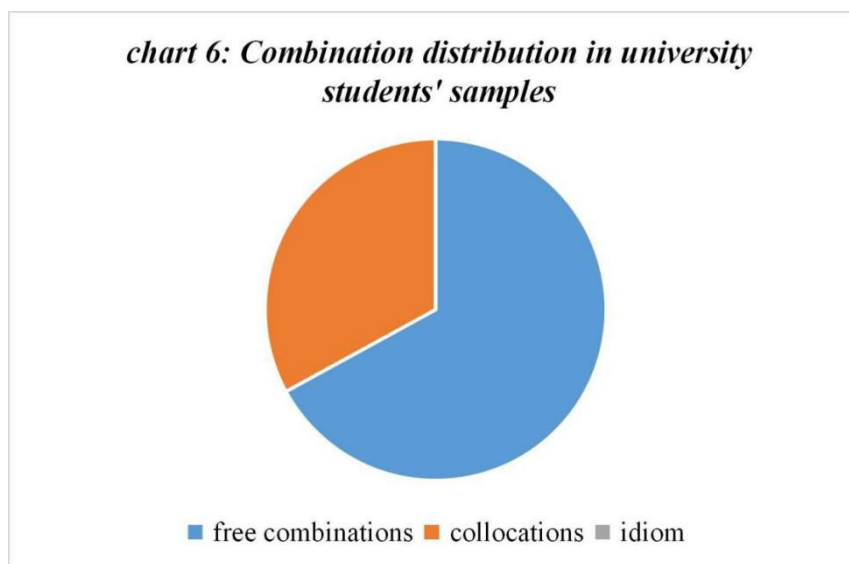
Figure 2. Combination Distribution in University Students' Samples

4.2 Analysis of Research Questions and Hypothesis

Research question 1 asked: to what extent do Chinese EFL learners use correct V-N collocations? As pointed out before, Nesselhauf (2003) assumed that clearly acceptable and largely acceptable combinations were correct while clearly unacceptable and largely unacceptable ones were wrong (Nesselhauf, 2003: 230). The criterion for judging collocations in the current study was based on that clearly acceptable and largely acceptable ones were considered correct while unacceptable ones were wrong. Altogether, 105 collocations were extracted and the distribution was presented in table 7.

Table 7. Distribution of Correct and Wrong Collocations

	Correct collocation	Wrong collocation	Total
Middle school students	37(79%)	10(21%)	47
University students	48(83%)	10(17%)	58

Notes: The percentage in the bracket indicates the proportion of correct/wrong collocations took up the total.

Thus, the answer was that middle school students produced 79% correct collocations of all their collocation production and university students did 83%.

Research question 2 asked: did collocation accuracy improve as overall L2 proficiency increases? The individual accuracy score was represented by the proportion of the correct collocations which took up the total. And independent T-test was conducted to compare the differences of the accuracy between the middle school and university students. This test was used because it could compare the means of each group and assess whether a significant difference existed on the accuracy between two groups of students. Table 8 and table 9 displayed descriptive statistics and results of independent T-test.

Table 8. Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Score	Middle school Students	10	.7700	.06377	.02017
	University students	10	.8190	.07680	.02429

Table 9. Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Score Equal variances assumed	.004	.953	-1.552	18	.138	-.04900	.03157	-.11532	.01732
Equal variances not assumed			-1.552	17.411	.139	-.04900	.03157	-.11548	.01748

As can be seen, the mean accuracy of university students (M = 0.82, SD = 0.77) was higher than that of middle school students (M = 0.77, SD = 0.64), but the improvement was not significant, t (17.41) = -1.55, p = 0.14, ns.. The results indicated that collocation accuracy would not improve significantly as the increase of overall L2 proficiency. Thus, the hypothesis

that Chinese EFL learners' collocation accuracy would improve with the increase of language proficiency has been proved to be invalid.

In order to clarify the hypothesis 3, questionable collocations were extracted and shown in table 10.

Table 10. Types of Mistake in Combinations and Their Distribution

Types of mistake	Example	Occurrence	
		Middle school students	University students
Verb Verb choice of non-existent verb, or wrong	*press the social burden (increase the social burden)	13 (42%)	16(46%)
Noun Noun choice of non-existent noun, or Wrong	*contain violent potent (contain violent content)	5(16%)	5(14%)
Preposition Preposition missing, present though unacceptable, or wrong	* influenced of our normal life (influenced our normal life)	4(13%)	3(9%)
Determiner Article or pronoun missing, present though unacceptable, or wrong	*lies in the behavior of public (lies in the behaviour of the public)	5(16%)	5(14%)
Usage 1 Combination exists but is not used correctly	*make order of the society (maintain social order)	0	2(6%)
Usage 2 Combination does not exist and cannot be corrected by exchanging or deleting single elements	*enhance our eyesight (broaden our horizon)	4(13%)	4(11%)
Total		31	35

As presented in table 10, improvements on various types of mistake were not equally distributed. Thus, the second hypothesis that improvement occurs in all types of mistakes proved to be invalid.

It could be found easily that there were obvious similarities between two level learners in L2 collocations production. First, types of mistake were almost consistent between two groups. Second, if different types of mistakes were ranked according to their frequency, the sequences of two groups were generally the same: wrong choice of verb on the top, then wrong choice of noun and usage 1 in the end. Therefore, hypothesis that similarities of errors would be identified between two groups of students was proved to be valid.

5. Discussion

5.1 Further Analysis of and Possible Explanations for the Results

First, consider the collocation mistakes. Results showed that errors in grammatical elements (including prepositions and determiners) accounted for 29% and 22% of middle school and university students' collocation production respectively. However, researchers (Gao and Zhang, 2015; Nesselhauf, 2003; 2005; Peng, 2016) suggested that mistakes in grammatical elements were less common compared with other error types. This indicated that lexical collocate such as 'acquire + knowledge' and 'miss + opportunity' might be easier to acquire than the whole collocations 'acquired the knowledge about' and 'miss an opportunity'. Therefore, instruction on the lexis only was never sufficient, and collocations should be taught as a whole. Another type of error should be noticed was the usage 2, which took up 13% and 11% of middle school and university students. These were the self-created combinations. Usage 2 suggested that learners might likely to use alternative expressions when having difficulties with unfamiliar collocations. Compared with students who only use collocations when they were sure those were correct (Biskup, 1992), Chinese EFL learners had a more positive attitude, and this should be encouraged.

It is worth mentioning that only considering the mistakes may lead to bias, so other characteristics should be noticed. Thus, both similarities and differences of collocation usage in detail will be discussed secondly. Despite the error types mentioned in the results, another similarity was that free combinations were found to be used most frequently among different levels of students, and then followed by collocations. However, no idioms were used by both groups. As shown in pie chart 4 and pie chart 5, free combinations accounted for approximately 60% of the whole combinations, which suggested that free combinations were easier to acquire. This was explained by scholars (Durrant and Schmitt, 2009; Nesselhauf, 2003; Peng, 2016) that free combinations were more creatively than strong associated collocations and idioms. However, it was the 'restricted expressions' that might improve the native-like language proficiency. Therefore, the results further supported that instructions on collocations was of great necessity and significance.

Thirdly, consider the accuracy. The results showed that the accuracy did not improve as the proficiency increase and the university students still made a lot of mistakes. This did not suggested that university students still had a low collocation competence. One possible explanation was that lower level students were likely to overuse a small quantity of collocations which they were familiar with, by contrast, high level learners may tend to use advanced collocations which were accompanied by more difficulties (Fan, 2009; Peng, 2016). In addition, 'the more difficult the newly acquired collocations are...the more mistakes higher level students will make in collocation production' (Gao and Zhang, 2005, p109).

Gao and Zhang (2005) also studied the co-efficiency of words (words used with high frequency and those with lower frequency) produced by different levels of learners and found that 'learners tend to use 'bigger' words with the increase of their L2 proficiency' (Gao and Zhang, 2005). They further suggested that the more errors advanced learners made, the more native-like competence they were achieving.

Two examples indicated that middle school students used simpler and fewer collocations than university students, but they had higher accuracy. One was 'V + knowledge'. 'Learn the knowledge' and 'get knowledge' were used by middle school students, while 'acquired knowledge', '*achieve knowledge' (*means a collocation which is largely unacceptable or unacceptable) and 'improve knowledge' were found in university students' samples. The other

example was 'V + problem'. Middle school students used 'have problems', by contrast, university students produced 'cause problem' 'solve problem' and 'meet problem'. Consistent with previous findings, results of the current study indicated that university students produced greater variety of collocations than middle school students, although their error rate was still high and some collocations were unacceptable in English.

5.2 Factors Affected Collocation Production

Various factors affecting collocation production can be identified. First, learners' L1 influence was found. The negative effect of L1 was revealed in misused incongruent collocations. Based on Wang (2011) and Yamashit and Jiang's (2010) study, congruent collocation was defined as collocations which could be directly translated word by word from L1 to L2, while incongruent collocation referred to those which could not be translated literally from its L1 counterpart. In the samples, incongruent collocations, such as 'make statement' and 'do homework' were wrongly used as 'do statement' and 'make homework'.

L1 influence was also observed frequently in the wrong choice of verb and noun, such as '*enhance the reading capability' (enhance the reading ability), '*results by the food' (caused by food) and '*stand for our school" (on behalf of our school). Such use might be explained by the fact that 'capability' and 'ability' can be translated into the same lexical item in Chinese, 'results' and 'cause', 'stand for' and 'on behalf of' are also in the same situation. This might be caused by students' disaquaintance with the difference between L1 and L2, and confusion with L2 lexis.

The second reason for the unacceptable collocation was the deficiency in L2 knowledge. Evidence were found in unacceptable collocation such as '*care for the news' and '*make responsibility for'. Students might know the collocations 'care about the news' and 'take responsibility for' and they attempted to express the intended meaning but encounter difficulties. Because they lacked lexis knowledge, and had confusions in 'care for' and 'care about' which resembled each other. They also could not accurately decide whether 'make' or 'take' was correct.

In sum, L1 influence and deficiency in L2 lexis were two main reasons for collocation errors.

5.3 Pedagogical Implications

Considering the two factors affecting the collocation production, we presented some teaching implications. Scholars have suggested that L2 collocation competence were related closely to L2 vocabulary proficiency (Fan, 2009; Nesselhauf, 2003; Yamashita and Jiang, 2010). One collocation cannot be acquired if its lexis was beyond students' existing capacity. Thus, teachers were expected to develop a wider vocabulary of learners. This was the first and foremost step in teaching collocations (Fan, 2009; Nesselhauf, 2003; Peng, 2016).

Second, in order to address the problem of negative L1 transfer, awareness of differences in collocation use between L1 and L2 should be increased. Lewis (2003) had proposed classroom activities for teaching collocations. In one of the activities, teachers present learners the possible words which can be collocated with the target word (such as achieve + success, goal, potential). After teaching them the detailed usage, teachers asked the students two questions: whether the same combinations are used in their L1, and are they used similarly or differently? By comparing L1 and L2 collocation use, learners noticed the differences between them and would be more sensitive to L2 collocation use. Thi s process was also consistent with noticing hypothesis which stated that learners acquire language by consciously identifying target language knowledge (Schmidt, 1994).

Furthermore, linguistics such as Laufer (2003) had found that the length of L2 exposure would reduce the L1 effect. Thus, students were expected to have sufficient amount of exposure to L2. Teachers could provide them with extensive reading and listing materials. Gao and Zhang (2015, p.111) suggested that good materials should contain 'collocations which are both undoubtedly acceptable and highly frequent in a neutral register'. Additionally, the L2 user corpora was definitely a useful source which affords a 'standard and plausible model' for L2 learner (Fan, 2009, p.121). Teachers could extract sentences which included target collocations by concordances, such as ConcApp (Greaves, 2005), and then edited them according to learners' proficiency. The materials can be organized as focus input especially to strengthen collocation competence. Besides, students could be given instructions on how to use corpus and provided with opportunities to do the corpus analysis, considering that the hardware device of Chinese classroom had been improved strikingly.

Third, the results showed that wrong choice of verb was the greatest problem for various levels of students. Therefore, attention should be paid especially on the choice of verb in collocation. However, focusing on the verbs did not signify to neglect other elements. As the results showed, Chinese learners also made a lot mistakes in non-lexical elements, such as the prepositions and articles. Thus, teachers should teach the whole collocations. As Kennedy (1990) had indicated, 'collocations are where grammar and vocabulary teaching meet', both grammatical and lexical elements should be taken into consideration in English class.

The present study presented evidence that collocations were significant in language development and language acquisition; supported that lexis should be granted a same position as grammar and confirmed that instruction should be included in English class. The results definitely supported Lewis's (2003) lexical approach which had been mentioned before. This paper also made contributions to Chinese EFL learners' collocation performance, and gained a better understanding of learners' collocation usage and difficulties.

6. Conclusion

This study explores Chinese EFL learners' V + N collocation production in practical usage. Accuracy was investigated, error types was identified, and pedagogical implications were discussed. The results showed that collocation accuracy will not improve with the development of L2 proficiency. Furthermore, it is L1 influence and L2 lexis deficiency that led to the mistakes. In addition, different levels of Chinese EFL learners exhibited both similarities and differences in V + N collocation usage.

The results made contributions to the collocation competence of Chinese EFL learners at different levels of proficiency. This study may be of interest to language researchers and language educators who have been exploring the learners collocation acquisition competence and researching methods to teach collocation.

Finally, limitations of the present research need to be addressed in the future. First, only 22 compositions were selected in this study, however, it would be better to include more samples. Second, only the V + N combinations were elicited and analyzed in the current study. Further researchers might concern other types of the combinations such as Adj + N.

References

Al-Zahrani, M. S. (1998). *Knowledge of English lexical collocations among male Saudi college students majoring in English at a Saudi university.* Ph.D. thesis, Indiana University of Pennsylvania.

Bahns, J. (1993). Lexical collocations: a contrastive view. *English Language Teaching Journal, 47*(1), 56-63. https://doi.org/10.1093/elt/47.1.56

Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System, 21*(1), 101-114. https://doi.org/10.1016/0346-251X(93)90010-E

Biskup, D. (1992). L1 influence on learners' renderings of English collocations: Polish/German empirical study. In: Arnaud, P.J.L. and Bejoint, H. (Eds.), *Vocabulary and Applied Linguistics*, 85-93. Basingstoke: Macmillan. https://doi.org/10.1007/978-1-349-12396-4

Boers, F., & Lindstromberg, S. (2009). *Optimizing a Lexical Approach to Instructed Second Language Acquisition.* UK: Palgrave Macmillan. https://doi.org/10.1016/j.system.2010.03.007

Cowie, A. P. (1998). *Phraseology: Theory, analysis, and applications.* Oxford: Oxford University Press. https://doi.org/10.2307/417696

Chen, P. C. (2002). *A corpus-based study of the collocational errors in the writings of the EFL learners in Taiwan*. Master thesis, National Taiwan Normal University. Https://Doi.Org/10.2307/417696

Chen, Y., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology, 14*(2), 30-49. https://doi.org/10.1111/j.1467-9922.2009.00559.x

Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics in Language Teaching, 47*(2), 157-177. https://doi.org/10.1515/iral.2009.007

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking and points of order. *Studies in Second Language Acquisition, 18*(1), 91-126. https://doi.org/10.1017/s0272263100014698

Ellis, N. C. (2001). Memory for language. In: Robinson, P. (Ed.), *Cognition and Second Language Instruction,* 33-68. Cambridge: Cambridge University Press. https://doi.org/10.1017/s0272263104243054

Fan, M. (2009). An exploratory study of collocational use by ESL students- A task based approach. *System, 37*(1), 110-123. https://doi.org/10.1016/j.system.2008.06.004

Gitsaki, C. (1996). *The development of the ESL collocational knowledge.* Ph. D. thesis, The University of Queensland.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: collocations and formulae.

In: Cowie, A. P. (Ed.), *Phraseology: Theory, Analysis and Applications,* 145-160. Oxford: Clarendon Press. https://doi.org/10.2307/417696

Harley, B. (1994). Appealing to Consciousness in the L2 Classroom. In: Hulstijin, J. H. and Schmidt, R. (Eds.), *Consciousness in Second Language Learning AILA Review.*

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics, 19*(1), 24-44. https://doi.org/10.1093/applin/19.1.24

Hsu, J. T. (2010). The effects of collocation instruction on the reading comprehension and vocabulary learning of Taiwanese college English majors. *The Asian EFL Journal, 12*(1), 47-87.

Hsu, L. C. (2005). *The effect of lexical collocation instruction on Taiwanese college EFL learners' listening comprehension*. Master thesis, National Kaohsiung First University of Science and Technology.

Hsu, J. Y. T., & Chiu, C. Y. (2008). Lexical collocations and their relation to speaking proficiency of college EFL learners in Taiwan. *The Asian EFL Journal, 10*(1), 181-204.

Hsu, J. Y. (2002). *Development of collocational proficiency in a workshop on English for general business purposes for Taiwanese college students*. Ph.D. thesis, Indiana University of Pennsylvania.

Hsu, J. Y. (2007). Lexical collocations and their relation to the online writing of Taiwanese college English majors and non-English majors. *Electronic Journal of Foreign Language Teaching, 4*(2), 192-209.

Kennedy, G. D. (1990). Collocations: where grammar and vocabulary teaching meet. In: Anivan, S. ed. Language Teaching Methodology for the Nineties. *Singapore: SEAMEO Regional Language Centre,* 215-229.

Laufer, B. (2003). The influence of second language on first language collocational knowledge and on first language lexical diversity in free written expression. In: Cook, V. (Ed.), *Effects of the Second Language on the First,* 19-31. Buffalo: Multilingual Matters. https://doi.org/10.1016/j.system.2004.12.004

Lewis, M. (1997). *Implementing the Lexical Approach*. London: Language Teaching Publication. https://doi.org/10.1002/j.1949-3533.1999.tb00159.x

Lewis, M. 1993. *The Lexical Approach: The state of ELT and a way forward.* London: Language Teaching Publications. https://doi.org/10.2307/3587574

Lewis, M. (2000). *Teaching collocation.* England: Language Teaching Publications. https://doi.org/10.1002/9781118784235.eelt0782

Lien, H. Y. (2003). *The effects of collocation instruction on the reading comprehension of Taiwanese college students.* Ph. D. thesis, Indiana University of Pennsylvania.

McCarthy, M. J. (1984). A new look at vocabulary in EFL. *Applied Linguistics, 5*(1), 12-22.

https://doi.org/10.1093/applin/5.1.12

McCarthy, M. (2004). *Touchstone: From corpus to course book.* Cambridge: Cambridge University Press.

Mounya, A. (2010). *Teaching lexical collocations to raise proficiency in foreign language writing.* Ph.D. thesis, Guelma University.

Movahediyan, A. E., & Allami, H. (2013). The effects of teaching lexical collocations on speaking ability of Iranian EFL learners'. *Theory and Practice in Language Studies, 3*(6), 1799-2591. https://doi.org/10.4304/tpls.3.6.1070-1079

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Newbury House. https://doi.org/10.2307/3587090

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Nation, I. S. P., & Newton, J. (1997). Teaching vocabulary. In: Coady, J. and Huckin, T. (Eds.), *Second language vocabulary acquisition*, 238-254. Cambridge: Cambridge University Press.

Nattinger, J. (1980). A lexical phrase grammar for ESL. *TESOL Quarterly, 14*(3), 337-344. https://doi.org/10.2307/3586598

Nattinger, J., & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press. https://doi.org/10.2307/329018.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics, 24*(2), 223-242. https://doi.org/10.1093/applin/24.2.223

Nesselhauf, N. (2005). *Collocations in a Learner Corpus.* Amsterdam/ Philadelphia: John Benjamins Publishing Company. https://doi.org/10.1017/S02722263107270068

Peng, X. L. (2016). *Use of Verb-Noun Collocations by Advanced Learners of Chinese*. Ph.D. thesis, University of Pennsylvania.

Richards, J. C., & Rodgers, T. S. (2014). *Approaches and Methods in Language Teaching.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511667305

Schmidt, R. (1993). Awareness and Second Language Acquisition. *Annual Review of Applied Linguistics, 1993*(13), 206-226. https://doi.org/10.1017/s0267190500002476

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review, 11*, 11-26.

Serrano, R., Stengers, H., & Housen, A. (2015). Acquisition of formulaic sequences in intensive and regular EFL programmes. *Language Teaching Research, 19*(1), 89-106. https://doi.org/10.1177/1362168814541748

Shin, D., & Nation, P. (2008). Beyond single words: The most frequent collocations in spoken English. *ELT Journal, 62*(4), 339-348. https://doi.org/10.1093/elt/ccm091

Sinclair, J. (1991). *Corpus, Concordance, Collocation.* New York, Oxford University Press.

Sung, J. (2003). *English lexical collocations and their relation to spoken fluency of adult non-native speakers.* Ph.D. thesis, Indiana University of Pennsylvania.

Tsai, K. (2015). Profiling the collocation use in ELT textbooks and learner writing. *Language Teaching Research, 19*(6), 723-740.

Tseng, F. P. (2002). *A study of the effects of collocation instruction on the collocational competence of junior high school students in Taiwan.* Master thesis, National Taiwan Normal University.

Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly, 44*(4), 647-668. https://doi.org/10.2307/27896758

Yazdandoost, Z., Amalsaleh, E., & Kafipour, R. (2014). The relationship among collocation knowledge and listening, speaking, reading and writing proficiency of Iranian EFL learners. *Journal of International Scientific Publications, 8*, 408-419.

Zhang, X. L. (1993). *English Collocations and their Effect on the Writing of Native and Non-native College Freshmen.* Ph.D. thesis, Indiana University of Pennsylvania.

Appendix A: Questionnaire for native speakers

1. Are your parents English native speakers?

2. Are you born in an English speaking country?

3. Did you grow up in an English speaking country?

4. Are you a students in Leeds University?

Appendix B: Instructions for native speakers

Instructions: You will be given 11 compositions of Chinese EFL learners (English as a foreign language). All the V-N combinations in the compositions are marked in red. Please judge the appropriateness of them. Please mark:

"C" for the correct ones;

"W" for incorrect ones;

"OK" for those you think are acceptable, but not the best way to say it;

"?" for those you cannot understand and cannot guess the meaning from the text.

For the "W" ones, please give your corrections; for the "OK" ones, please give a better expressions if possible. The judgement take into account the combinations in their entirety

(including, for example, the modification and complementation pattern of the nouns). Please consider the errors like verb choices, noun choices, prepositions, complements and pragmatic errors. And please complete those that you think are incomplete. "//" separates two collocations, for example: "arrange // and take part in some school activities" are considered as two combinations "arrange some school activities" and "take part in some school activities".

Copyright Disclaimer