

Testing Korean EFL Learners' Verb-particle Collocation Knowledge and Use

Hyojin Jeon

(Seoul National University)

Jeon, Hyojin. 2011. Testing Korean EFL Learners' Verb-particle Collocation Knowledge and Use. *The Linguistic Association of Korea Journal*. 19(3). 25-46. This paper aims to assess EFL learners' knowledge of English verb-particle collocations via general contexts in English. The "English Verb-Particle Collocation Test (VPC test)" was developed and administered to 134 Korean EFL learners at the university level. The VPC test consists of three subtests divided by test item types in line with recognition, recall, and production: a multiple-choice test, a gap-filling test, and a writing test. All the data were analyzed using statistical and psychometrical methods. The results of analysis revealed that the VPC test achieved not only an acceptably high degree of test score reliability for the two multiple-choice test sections, but also a solid reliability and a high correlation with the *TOEFL (convt)* scores. This study supports the argument that the VPC test created for the study is a reasonable measure of Korean EFL learners' collocation knowledge. The study also demonstrates that test item formats and test length have a considerable impact on the reliability and validity of test scores. One more important finding was that the examinees tended to use specific, identifiable types of VPCs according to their essay test scores. Finally, this study reveals that VPCs can be indirect indicators of proficiency in EFL education.

Key Words: verb-particle, multi-word verbs, verb collocation, second language assessment, language test

1. Introduction

Much research has increasingly revealed that collocations have a positive

relationship with learners' proficiency, especially in the fluency and accuracy of the target language (Bahns & Eldaw, 1993; Benson et al., 1997; Bonk, 2001; Ellis, 2001; Howarth, 1998; Lewis, 1993; Nasselhauf, 2005; Read, 1993, 1998; Runcie, 2002; Schmitt, 2004; Skehan, 1998; Wray, 2000). In English, verbs are the head of phrases or sentences, and L2 learners often have trouble using verbs correctly in both word choice and order. Verb-embedded collocations indeed make up an important part of English collocations. Collocations can be a good medium to reflect the complexities of lexis, in that collocations integrate comprehensive use of words in a certain context in relation with other words. Therefore, this study primarily focuses on testing EFL learners' knowledge and use of English verb-particle collocations (henceforth, VPCs)¹⁾, and considers different cognitive levels of language learners—recognition, recall, and production—as another fundamental issue, while incorporating the effects of contextual information surrounding the target lexical items is of secondary importance. To achieve the main goals, this study poses three research questions below.

1. What relationships exist among the subtests of the VPC test and between the subtests and measures of overall English proficiency?
2. Are there any verb-particle collocation types or item types that cause particular difficulty for the construction of a reliable VPC measure?
3. Are there some significant patterns of EFL (English as a foreign language) learners' VPC use in their English production?

To answer the questions and put forward evidence to support the analyses, adult Korean EFL learners at the university level or higher were examined, who possibly have more exposure to English than those who completed their formal education in Korea; it seems plausible that learners who study for a longer period of time will generally have more opportunities to improve VPC knowledge and use. This study develops an "English Verb-Particle Collocation Test" targeting Korean EFL learners, investigates the test's reliability and validity, and initiates a discussion related to test item types and EFL learner

1) Verb-particle collocations are synonymous with multi-word verbs. A multi-word verb refers to a verb that consists of a basic verb and another word or words (preposition and/or adverb). This study adopts "verb-particle collocation" as a cover term.

performance on language tests by multiple methods, such as multiple-choice tests, gap-filling tests, and essay writing tests.

2. Background

2.1. L2 learners and the Significance of Verb-particle Collocations

Using collocations implies the speaker has an enriched vocabulary. In other words, using correct collocations more closely resembles language production of native speakers; using wrong collocations is an error. VPCs can be roughly classified into three subgroups, depending on whether the particle is an adverb or preposition: phrasal verbs (e.g., *get up*), prepositional verbs (e.g., *believe in*), and phrasal-prepositional verbs (e.g., *get on with*). The types of VPCs are particularly restricted to the constructions of verbs and particles linked with them: for example, VPCs are often classified according to grammar, such as intransitive (e.g., *come to*), transitive inseparable (e.g., *look into*), and transitive separable (e.g., *put out*). Although some VPCs cannot be replaced by one-word verbs in some cases—for example, *make up*, the function of VPCs is almost the same as that of a single verb in the syntactic and lexical contexts.

It is difficult for EFL learners to have some inkling of the precise meanings of a significant number of VPCs. In many cases, the VPC meanings are not clearly caught by the sum of the meanings of its elements; they are somewhat idiomatic, and occasionally polysemous. This leads the learners to feel frustrated, for there is no clear-cut way to easily distinguish their lexical meanings from the original meanings of their parts. Thus, it is acceptable that L2 learners are known to have a tendency to prefer one-word verbs, due to the unfamiliarity and lack of confidence in using those multi-word verbs (Dagut and Laufer, 1985; Granger, 1998; Schachter, 1974; Siyanova & Schmitt, 2007).

In contrast to L2 learners' difficulties with using VPCs, native speakers widely use VPCs in their daily life. Accordingly, VPCs are viewed as being colloquial and as an important component of spoken or informal English (Altenberg, 1998; Biber et al., 1999; Dagut & Laufer, 1985; Siyanova & Schmitt, 2007). However, VPCs are not limited solely to spoken or informal language in

real-life. Even though learners' L2 use might be mostly confined to the classroom settings, it is true that the use of VPCs can abound within out-of-classroom settings: consulting with a department manager, meeting with project-team members, ordering in the campus cafeteria, et cetera. Therefore, it has to be said that EFL learners' use of VPCs can be connected with the nearly every aspect of their English-speaking context.

2.2. Considerations for Test Development

2.2.1. Test Design, Reliability, and Validity

In language testing and assessment, numerous models and frameworks can be used because there are a variety of targets, such as language proficiency, or communicative competence (communicative language ability). According to Bachman and Palmer's (1996) framework, which is one of the most significant and influential to date, the first step in language test design is to define the purpose of the test. Tests can be used for various purposes such as diagnosis, placement, assessing proficiency, and measuring achievement. It is important to clarify the intended purpose of the test because a well-made test helps testers to make relevant decisions about learners or language programs (Read, 2000).

Once the purpose has been defined, the primary focus becomes demonstrating how reliable the test is and how valid the interpretations and use of test scores are. A test can be both reliable and valid, one or the other, or neither. The main concerns of reliability and validity are complementary in designing and developing tests; the former is for minimizing the effects of measurement error, and the latter is to maximize the measures of test takers' language abilities via the tests (Alderson et al, 1995). According to Worthen (1993), reliability is the measure of how stable, dependable, trustworthy, and consistent a test is in measuring the same thing multiple times. Validity is dependent on the test's content and method, in addition to how test takers perform (Bachman, 1990). The degree of validity of a test basically depends on the test's purpose; in practice, validity cannot be achieved without reliability. However, the problem is that maximizing the reliability often requires reducing validity; this is the reason why Alderson et al. (1995) argued that there is a trade-off between the two degrees. Therefore, reliability is a prerequisite of

validity of a test (Bachman, 1990), while validity is a fundamental component that has to do with evaluating whether a test measures what it is supposed to. It is inevitable that a test should be checked in as many ways as possible, which include various components of reliability and validity.

2.2.2. Recognition, Recall, and Production in Language Testing

Test takers usually understand the meaning of target words they are presented with, and are able to elicit the target word forms from their memory; the former is called "recognition," and the latter is "recall." Read (2000) pointed out that recognition and recall in relation to vocabulary tests can be assessed by selective and relative context-independent test items. Multiple-choice test items are widely used for recognition tests, while translation or gap-filling items are generally used for recall tests in language testing. Test takers activate different problem-solving skills and cognitive processing levels in their memory to answer each type of task. In this regard, one of the most demanding cognitive processing levels can be "production," since it requires the "higher-order executive process that provides a cognitive management function in language use" (Bachman, 1990, p.70). Free composition and oral interview are traditional tests of general language ability in relation to production. With regards to vocabulary, production tests have examined the depth of understanding of vocabulary elements, such as collocational knowledge, and precision in lexicon (Bogaards, 2000; Henriksen, 1999; Quian and Schedl, 2004; Wesche and Paribakht, 1996).

3. Method

3.1. Participants

The participants for the study were five native English speakers and 134 Korean undergraduate students, all studying at a university located in Seoul, Korea. However, the native English speakers were not included in the examinee group for data analysis. Instead, their answers were reviewed for the purpose of seeing native speakers' tendencies and to make sample answers to compare with

those of Korean participants. Their ages were between 21 and 45; they came from the United States, Canada, and New Zealand. Three of them were exchange students and the others were native-speaking English instructors.

The 134 Korean participants were comprised of 58 females and 78 males. All of them learned English after their native language (L1) was fully acquired. Their mean age was 24.56, and the average number of years spent living in countries where English is spoken as either a native or an official language was 2.82 years²⁾. Their majors were widely distributed over 30 different fields of study, but none of them were English majors. Their general scores from TOEFL and TEPS tests were interpreted as being above high-intermediate level³⁾ in their English proficiency. However, the results according to the age, length of foreign residence, major, and even their TOEFL and TEPS scores did not make a meaningful difference in the VPC test scores, and thus the participants were not grouped for the study by those factors.

3.2. Instruments

3.2.1. Preliminary

Two different pilot versions were administered in the preliminary study. In the first phase of the preliminary study, 32 undergraduate students of non-English majors were voluntarily recruited from SNU (Seoul National University). They all had advanced levels of English communicative competence. A 26-item questionnaire of a six-point-scaled "Verb-pair questionnaire"⁴⁾ was administered to the participants. The results showed that their tendency differed from that of the non-native English speakers in Siyanova and Schmitt's study (2007); the L2 learners of Siyanova and Schmitt (2007) had a native-like sense of phrasal verbs in spite of a slight preference for a few items.

In the second phase, another group of seventeen undergraduate and

2) The number of participants and their length of residence are as follows: 0~0.5 years=84, 0.6~1 years=3, 1.1~2 years=17, 2.1~5 years=0, and more than 6 years=30.

3) Dividing the participants by their language proficiency level followed the criteria suggested by the TEPS guideline for score interpretation. TEPS regards the test takers whose scores are between 600~700 as high-intermediate (Grade 2), between 700~800 as advanced (Grade 2+), and over 800 as near-native level (Grade 1).

4) It was exactly the same questionnaire created by Siyanova and Schmitt (2007).

graduate students of advanced levels were recruited. All of them were also non-English majors. For testing material, a "VPC test"⁵⁾ was created and provided to the participants, which was done in order to sort out the testable VPC test items having moderate item difficulty. The study found that, even though the participants were able to recognize VPCs and distinguish between particles and prepositions, they were unable to use them properly.

3.2.2. Verb-particle Collocation Test

After evaluating the results of the two pilot studies, the study took further steps to examine the relationship between language proficiency and knowledge of VPCs, and the effects of contextual information on learners' understanding and use of VPCs. For this purpose, a new version of the "English VPC test" was created. This new VPC test used in the main study was divided into three subtests, which were designed to assess the examinee's collocation knowledge at three different levels: recognition, recall, and production. Table 1 shows the overall structure of the total test and subtests as well as VPC types included in the subtests.

Table 1. The Breakdown of Each Section in the VPC test

| Sub-division | Task type (Item type) | Target patterns | Item No. | N |
|--------------|--------------------------|-----------------------|----------|----|
| Section A | Rcg (MC) | Vi + ptcl | 1-3 | 3 |
| | | Vi + prep | 4-6 | 3 |
| | | Vt + immovable ptcl | 7-9 | 3 |
| | | Vt + movable ptcl | 10-11 | 3 |
| | | Vt + prep | 12-15 | 3 |
| | | Vt + noun + prep | 16-18 | 3 |
| | | Vt + prep phrase | 19-21 | 3 |
| | | Total | | 21 |
| Section B | Rcl (GF) | Verb matching | Part 1 | 5 |
| | Rcg (MC) | prep /ptcl matching | Part 2 | 5 |
| | | Total | | 10 |
| Section C | Prd (W) | Impromptu composition | | 1 |

Note. No.=number, N=number of the items; ptcl=particle, prep=preposition, Rcg=Recognition, Rcl=Recall, Prd=Production, MC=Multiple-choice, GF = Gap-filling, and W=Writing

- 5) The test items were modified from sentences which appeared in the exercises in *Oxford Collocations Dictionary for Students of English* (Oxford University Press, 2002) into multiple-choice test items.

The first subtest, Section A, consisted of 21 multiple-choice items that were intended to assess VPC knowledge in the limited context of a single sentence, containing all types⁶⁾ of VPCs evenly. The items had been previously trialed in the second phase of the preliminary study; they were selected and revised after eliminating extremely easy and difficult ones, based on the item facility index in the preliminary study.

The second part of the test, Section B, was designed to test whether the learners possessed abilities to deal with somewhat more complex and demanding tasks. Divided into two parts, Section B included 10 gap-filling items with extended contextual information through reading passages.⁷⁾ In the first part, test-takers were required to fill in the blank of the questions with verbs or VPCs that appeared in the reading passage. The second part required test-takers to read a passage in the form of a newspaper clipping, and then fill in blanks in the following passage, which took the form of a letter to a close friend; this is analogous to a storytelling task, which could convey given contexts from formal language to colloquial, or informal, language.

Section C asked the test-takers to write an impromptu essay in a letter form. This section was to see the degree of Korean EFL learners' VPC use. A constructed-response task required the test-takers to make full use of their knowledge, experience, and imagination after reading a short prompt. The quantity of VPCs and the quality of writing for Section C were both examined to find some characteristics or relationships between their writing test scores and English proficiency. Thus, the test takers' written production was thoroughly investigated from various points, such as descriptive statistics, the VPC pattern analysis, and error analysis.

6) This study reclassifies VPCs into seven units, according to the patterns of the associated words: (1) intransitive verb plus particle (2) intransitive verb plus preposition (3) transitive verb plus immovable (i.e., inseparable) particle (4) transitive verb plus movable (i.e., separable) particle (5) transitive verb plus preposition (6) transitive verb plus noun and preposition and (7) transitive verb plus prepositional phrase. The distinction among grammatical, lexical, and idiomatic collocations is thereby no longer necessary.

7) The reading passages of Section B in the test were extracted from *Really Learn 100 Phrasal Verbs* (Oxford University Press, 2004) and *Really Learn 100 More Phrasal Verbs* (Oxford University Press, 2007).

3.2.3. Questionnaire and Follow-up Interview

A questionnaire consisting of 15 items was administered to the test-takers to collect information on the factors related to the learner variables. This was designed to collect feedback about the test itself, and investigate their preferred styles and ideas on learning as well as their extra-linguistic backgrounds, such as age, gender, major, et cetera. Except for short-answer items, all of the items were typical Likert-type questions based on a 5-point response scale. Participants were also asked to provide their TOEFL and/or TEPS scores in the questionnaire. As the raters scored their performance on the VPC test, interviews were done with 17 participants whose scores or writings were in peculiar patterns.

3.3. Procedures

The participants were recruited by posting an advertisement of the research on the university web pages and bulletin boards. The VPC test, questionnaire, and interview were conducted mainly in their classrooms or on campus by appointment. The participants were given 50 minutes to finish the whole test and answer the questionnaire. The researcher separately recorded the time required for all data collection; on average, 45.13 minutes was spent both to complete the test and answer the questionnaires.

Both Sections A and B of the VPC test were scored by the researcher. A correct response by a participant in Sections A and B was recorded as 1, and an incorrect response as 0. Still, Section B admitted partially correct answers as well, giving 0.5 in cases where the test-takers answered with the correct target word but in the wrong form. For Section C, three raters assigned holistic scores, ranging from 0 to 5, based on the scoring rubric of the independent writing section of TOEFL *iBT*; the independent writing task for this study is relatively closer to personal or casual situations than the independent writing task of TOEFL. To grade the essays for Section C fairly, two more raters⁸⁾ were

8) One external researcher was an incumbent interpreter working for an international broadcasting company. The other was a teaching assistant with a near-native level of English proficiency, who graduated from a graduate school of interpretation and translation and worked as an English teacher in public middle schools.

employed for scoring the essays. Since the three raters needed to adjust the manner and rules somewhat to match their own interpretations, they studied and reviewed the TOEFL scoring rubric carefully, and then marked some selected students' essays together and resolved score discrepancies through discussion.

Microsoft Excel 2007 was used to process the scored items response data, compute simple statistics, and create charts for the raw scores of the test items. For more complex statistical and psychometric analyses, Statistical Package for the Social Sciences (SPSS; Version 12.0) was used in this study.

4. Results

4.1. Descriptive Statistics

All participants have one or more scores from internationally recognized English proficiency tests, and the majority of the participants (89 out of 134) had computer-based or internet-based TOEFL (TOEFL *iBT*) scores. Hence, non-TOEFL *iBT* scores were converted into TOEFL *iBT* scores according to the score conversion table provided by ETS and LEI at SNU. Table 2 displays the descriptive statistics.

Table 2. Descriptive Statistics for the Scores from the VPC test and TOEFL (*convt*)

| | | M | SD | SR | Min | Max |
|------------------------|-----------|-----------|------|------|-----------|----------|
| Section A | | 13.1 (62) | 4.2 | 16 | 5 (24) | 21 (100) |
| Section B | Part 1 | 4.07 (81) | 0.82 | 4 | 1 (20) | 5 (100) |
| | Part 2 | 3.39 (68) | 1.16 | 4 | 1 (20) | 5 (100) |
| | Sub-total | 7.5 (75) | 1.5 | 7 | 3 (30) | 10 (100) |
| Section A+B | | 20.6 (66) | 5.1 | 19 | 11 (35) | 30 (97) |
| Section C | | 3.4 (68) | 0.7 | 4 | 1 (14) | 5 (100) |
| VPC-test Total | | 24 (67) | 5.4 | 20.5 | 13.5 (38) | 34 (94) |
| TOEFL (<i>convt</i>) | | 90.6 (76) | 11.7 | 47 | 59 (58) | 116 (97) |

Note. Mean=the average scores, SD=Standard deviation, SR=Score range, Min=Minimum score, Max=Maximum score, ()=grade on a scale of one hundred points; TOEFL (*convt*)=Converted scores of all English proficiency test scores the participants had on the TOEFL *iBT* basis

Judging from the means and perfect scores, Section A had the lowest rate of correct answers (62); on the other hand, participants gave the highest (75) in Section B. This can roughly be interpreted as meaning that Section A was the most difficult section in the test, whereas Section B was the easiest.

However, Section B has its own subsections as well, Parts 1 and 2. Part 1 recorded the highest mean score in terms of the score (81), whereas Part 2 achieved a mean score (3.39, 68) that is similar to other sections of the VPC test and the total test. Meanwhile, when the means of the VPC test and TOEFL (*convt*) were compared in reference to the highest possible score of the respective tests, it turned out that the participants on average scored higher on their TOEFL (*convt*) (76) than the VPC test (67) in terms of percentile score. Of course, this should be interpreted with caution since the TOEFL *iBT* scores are scaled scores whereas the VPC test scores are percent correct scores.

4.2. Reliability

4.2.1. Internal Consistency and Inter-rater Reliability

This section adopts internal consistency to see if the test takers responded to the test items consistently in a single trial. The reliability coefficients were 0.78 for Section A, 0.45 for Section B, and 0.80 for the two sections combined. When these were adjusted for standardized items, their values were slightly changed to 0.78, 0.51, and 0.79, respectively. These reliabilities marked by Cronbach's alpha represent the consistency of test scores; the alpha is expected to increase when more items are added to the test.

With respect to inter-rater reliability, which is estimated in order to minimize variation and increase validity in the application of a scoring rubric. A series of *t*-tests was conducted to test the statistical significance of the differences in the rating means among the three raters. According to the results of *t*-tests, the rating mean differences were statistically significant between Raters 1 and 2 ($t=-3.24$, $p=0.002$) and between Raters 1 and 3 ($t=-3.27$, $p=0.001$). In contrast, the rating mean difference between Raters 2 and 3 was not statistically significant, $t=1.00$, $p=3.19$. This means that Rater 1 can be regarded as a stringent rater while Rater 2 is lenient. Rater 3 is about average between the other two raters.

4.2.2. Correlations among Subtests and Other Criterion Measure

Table 3 provides the observed and disattenuated correlations among the VPC subtest scores, the VPC total test score, and an external criterion measure, TOEFL (*convt*). The values presented in the parentheses in the table are the disattenuated correlations.

Table 3. Pearson Product-moment Correlations for the VPC test and TOEFL(*convt*)

| Section | A | B | A+B | C | A+B+C | TOEFL (<i>convt</i>) |
|---------|--------------------------|--------------------------|--------------------------|-----------|-------|---------------------------|
| A | 1.00 | | | | | .81 (.94) |
| B | .43 (.71) | 1.00 | | | | .59 (.90) |
| A+B | .95 (1.00 [*]) | .69 (1.00 [*]) | 1.00 | | | .85 (.98) |
| C | .29 (.33) | .18 (.21) | .29 (.34) | 1.00 | | .30 (.32) |
| A+B+C | .94 (1.00 [*]) | .68 (.72) | .99 (1.00 [*]) | .41 (.44) | 1.00 | .85 (.91) |

Note. All significant at .01 level Cronbach's alpha= .81, Cronbach's alpha on standardized items= .91; ()=disattenuated correlation coefficient; ^{*}=disattenuated values greater than 1.00⁹⁾

As shown in the table, Section A had a high correlation with the TOEFL *iBT* scores ($r=0.81$), a moderate level of correlation with Section B ($r=0.43$), and a weak level of correlation with Section C ($r=0.29$). It is noteworthy that Section C also has rather low correlations with other tests: 0.29 with Section A; 0.18 with Section B; 0.29 with the combined Section A plus B, and; 0.30 with TOEFL (*convt*). This is possibly caused by the different test characteristic of Section C; as explained in the previous section, Section C was a writing test having a single item, whereas Sections A and B consisted of more than at least ten test items. Another possibility is that Section C has a rather limited score range of the examinees from the lowest score to the full score; the numerical values from 0 to 5 stand for grade by scoring rubric. Figure 1 presents a scatter plot that shows the strength of association between the VPC total test and TOEFL (*convt*)

9) Since each section employed different test formats, disattenuation was done in connection with calculating the observed. Disattenuated correlations are the numerical values in parentheses in Table 3. Four values (1.00^{*}) exceeded the perfect correlations after disattenuation. The values between Section A and Section B indicate that the observed correlation of them was dramatically improved: 0.43 to 0.71. Section C left something to be desired, as its correlation coefficients increased marginally more than the observed values: 0.41 to 0.44.

scores, which was drawn to provide a clearer picture of the relationship between the two test scores.

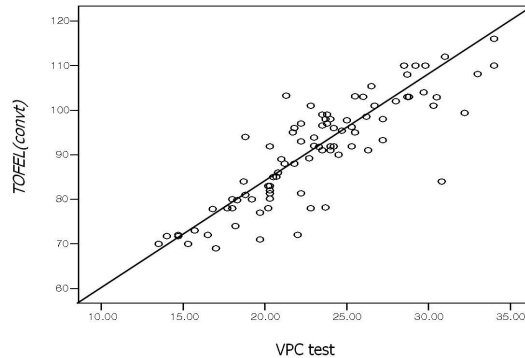


Figure 1. Scatter Plot of the VPC Test and TOFEL Scores

4.3. VPC Analyses

As preparatory investigation, the patterns of the test-takers' use of verb-particle collocations in the essays were examined. Table 4 offers a brief descriptive overview of the total written data collection. The average length of essays was 15.1 sentences, with 197.5 words in total, and the total number of VPCs per essay ranged from 0 to 11, with a mean of 3.5. The type-token ratio of VPCs was 2.66; this means that the data contained 823 VPC tokens (total number of times a VPC item is used) and 309 VPC types (the actual number of the VPC used) on the lemma basis¹⁰⁾.

Table 4. VPCs Appeared in Test-takers' Essays for Section C

| Frequency unit | Mean | Total |
|--------------------------|-------|---------|
| Words | 197.5 | 26,465 |
| Sentences | 15.1 | 2,023.4 |
| VPC tokens | 6.14 | 823 |
| VPC types | 3.5 | 309 |
| Type-token ratio of VPCs | 1.75 | 2.66 |

10) Lemma is a canonical form of a word considered as its citation form together with all the inflected forms—regardless of the tense, number, or voice. For example, the lemma *go* consists of *goes*, *going*, *went*, and *gone*.

Table 5 gives the most frequently-used VPC item lists and their frequencies. The total number of Type 1 verb particles used in student essays was 355, with the five most frequent verb particles being: *break out*, *go out*, and *come back*. Also, the total number of Type 4 used was 346, with the most frequent verb particles being *find out*, *put out*, and *turn out*.

Table 5. The List of Most Frequently-used VPCs
belonging to Types 1 and 4

| Total | f | % | Type 1 | f | % | Type 4 | f | % |
|-----------------------|----|-------|-----------|----|-------|----------|----|------|
| break out (Type 1) | 55 | 17.85 | break out | 55 | 17.85 | find out | 32 | 10.2 |
| find out (Type 4) | 32 | 10.2 | go out | 36 | 5.74 | put out | 32 | 10.2 |
| put out (Type 4) | 32 | 10.2 | come back | 24 | 3.83 | burn out | 30 | 9.56 |
| turn out (Type 4) | 30 | 9.56 | wake up | 24 | 3.83 | turn out | 30 | 9.56 |
| burn out (Type 4) | 30 | 9.56 | go on | 16 | 2.56 | blow out | 30 | 9.56 |

Note. f=frequency of each VPC, %=percentage of each VPC out of the total

The VPCs used in test taker's essays were broken down according to the test-takers preference of VPC types for in-depth examination, and then arranged by their essay scores; these are displayed in the bar graphs in Figure 2.

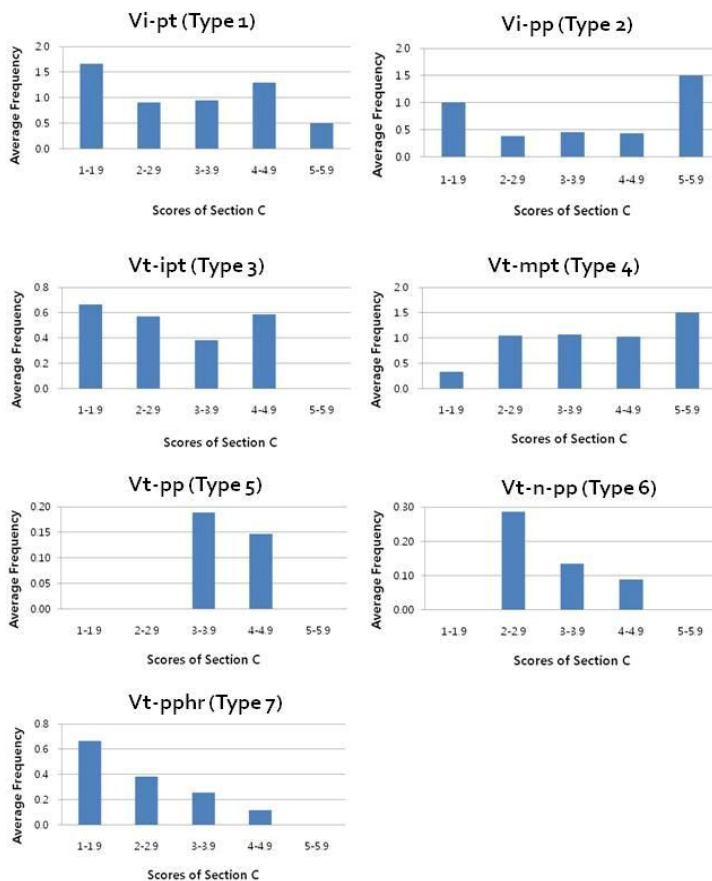


Figure 2. Averaged VPC frequencies for each essay score level

The results suggest that Korean EFL learners are not accustomed to VPCs belonging to Types 5, 6, and 7, and that they might have more exposure to Types 1 to 4. This tendency is strongly in line with the patterns of test takers' VPC use. However, Figure 2 presents intriguing trends in two different patterns of examinees' VPC use. First of all, salient decreasing trends are captured; the use of VPCs belonging to Types 1, 6, and 7 tends downward. As seen, VPC groups of Types 6 and 7 appear infrequently in examinees' writings. It is therefore more important to pay attention to the learner tendency in using Type 1 VPCs compared with Type 4 VPCs, in that Types 1 and 4 were the most common VPC types used and had similar frequency rates. At the same time,

though, examinees' use of Type 1 VPCs exhibits the opposite tendency in their use of Type 4 VPCs; as learners have higher writing proficiency, they seem to avoid using VPCs belonging to Type 1.

On the other hand, there is an increasing trend in the use of Type 4 VPCs. As the writing test scores for Section C go up, the frequency of Type 4 increased. This means that learners who have more knowledge and higher proficiency levels may favor VPCs belonging to Type 4 and feel confident of using Type 4 VPCs. Since the essay scores are generally proportional to examinees' language proficiency levels, this suggests that low-level learners might have less knowledge of Type 4 VPCs than high-ability learners, and avoid producing Type 4 VPCs.

5. Discussion

5.1. Relationships among the Subtests, Total Test, and an Overall Proficiency Measure

This section attempts to answer the first research question: what relationships exist among the subtests of the VPC test and between the VPC test and an external overall English proficiency measure (TOEFL *iBT*). When the disattenuated correlations were used for comparison, both Sections A and B were highly correlated. These two section scores also had very high correlations with the converted TOEFL *iBT* score. This suggests that the combined section score of two multiple-choice test of Sections A and B can serve as good predictors of the overall English proficiency represented by the TOEFL *iBT* score.

On the other hand, Section C had very low correlations with the other two VPC sections and an external criterion measure, TOEFL (*convt*), despite a high level of inter-rater reliability achieved for this section. This means that the nature of the assessment construct being measured by Section C might be quite different from those measured by other sections of the VPC test used in the study. Another possibility is that the type of a rating rubric used in this study to rate the test-takers' essays might have some impact on the pattern of

relationship among the VPC subtests. The holistic scoring rubric for TOEFL *iBT* independent writing tasks used in this study might not have been sensitive enough to capture the test-takers' ability to appropriately use the VPCs in their essays. One interesting question in this regard is whether the observed pattern of relationships among the essay section and other sections of the VPC test could have been very different, if a multi-trait scoring rubric or other lexical measures of evaluation were used to assess the accuracy and appropriateness of the test-takers' use of VPCs in the essays. Such examples of lexical measures may include lexical variety and accuracy measures, the frequency of various types of VPCs, and the frequency of errors associated with the use of the collocations in the test-takers' essays.

5.2. VPC frequency and patterns in the essay test

The second and last research question for this study was to ask whether there are some source of difficulty and significant patterns of Korean EFL learners' VPC use in their English production. One intriguing thing was found when analyzing the actual VPC scope of the participants' compositions. The frequencies of each VPC type used in the test-takers' essays were computed and ranked. Interestingly, the most frequently used VPC (*break out*) turned out to be the one that also appeared in the essay prompt for the writing test of Section C. This may reflect a typical examinee test-taking strategy of copy and paste from stimulus materials, which the test-takers utilize to compensate for a lack of lexical and grammatical knowledge required to complete the writing task at hand. Again, this emphasizes the importance of minimizing the possibility of the examinees' copying from input materials and finding a way to penalize plagiarism in EFL writing assessments.

One more finding deserving mention can also be found in the examinees' VPC using patterns. VPC Types 1, 2, 3, 4 (Vi+particle, Vi+preposition, Vt+immovable particle, and Vt+movable particle, respectively) are preferred in Korean EFL learners VPC use in general, and VPC Types 5, 6, 7 (Vt+preposition, Vt+noun+preposition, and Vt+ prepositional phrase, respectively) are much less used. One probable assumption is that the former includes more familiar types of VPC items to Korean EFL learners and the latter does not; or perhaps, the

former is more frequently taught in Korean EFL settings relatively than the latter. Another interesting pattern was that Types 1, 6, and 7 displayed decreasing patterns as the scores go up. Further investigation is warranted to ascertain whether the above assumption is supported.

The two most frequently used VPC pattern types, Types 1 (Vi+particle) and 4 (Vt+movable particle), showed contrasting tendencies in VPC use patterns while Type 1 showed a decreasing pattern in usage, Type 4 showed an increasing pattern. In practice of EFL teaching and learning, Type 1 is taught and used as a single lexical item from their earlier stages of EFL learning, which might be regarded as elementary level VPC Types; in addition, most of them represent simple action or status, which are also used in many common situations in real-life. By contrast, VPCs belonging to Type 4 are taught in later stages in EFL educational convention, in that Type 4 requires more knowledge than Type 1: for example, the necessity of the objective and the movability of the particle. Consequently, Type 1 is more accessible for learners to use commonly in informal and spoken discourse, while Type 4 VPCs are more high-dimensional to use in daily life and thus used in more formal and written discourse. If this is true, it produces counterevidence to a generally accepted notion that VPCs are regarded as colloquial, which are an important component of spoken or informal English (Altenberg, 1998; Biber et al., 1999; Dagut & Laufer, 1985; Siyanova & Schmitt, 2007). At the same time, it also supports the idea that substantive use of specific lexical items can help identify the examinee parameters. This can be the ground for the idea that learners' VPC knowledge and use are decent indirect indicators of EFL learners' language proficiency level. Of course, to confirm and clarify this point, a follow-up study concentrating on Types 1 and 4 should be done in the near future.

6. Conclusion

This study explored the knowledge and use of VPCs of Korean EFL learners. In this study, the notion of VPCs was defined as a category of verb collocations in English, which were again grouped into the seven types on the basis of grammatical features. Also, an "English verb-particle collocation test (VPC test)"

was created and administered to Korean EFL learners at university levels. Judging from the test score reliability and criterion-related validity evidence, it can be said that the VPC test developed for this study is a reasonable measure of Korean EFL students' collocation knowledge as well as overall English language proficiency. This study also demonstrates that an intensive rater-training is required to minimize the rater effects on scoring. Furthermore, it is noteworthy that the participants' performances on each section were somewhat different, which may demonstrate the effects of test methods on test scores obtained through these measures. It would be an important matter of language testers to find a way to minimize the impact of such construct-irrelevant factors on examinees' test scores.

The findings of this study offer a reason as to why researchers in this field should pay more careful attention to investigating the potential consequences of employing certain types of test formats in their tests in terms of score interpretation in later stages. This should not necessarily restrain language test developers from trying out a variety of item types and formats in their tests. What is really important is that the effects of these items should be carefully studied and examined.

Another finding in this study offers significant evidence that VPCs, such as Type 1 and 4, can be indirect indicators in L2 teaching and learning. In other words, assessing certain types of VPCs having diagnostic values helps EFL teachers and learners to find the learners' language proficiency level. The study gives potential clues that the degree of certain VPC use may result from the different cognitive processing levels. Although using contextual information in the given discourse helps to determine the choice of formal/informal and written/spoken register, this can be part of the process of corresponding lexical items in language users' mind.

Although this study sheds some important light on the testability of VPC knowledge for Korean EFL learners, several limitations of the studies were identified with respect to the test design and research methods used for the study. The first one is an unequal numbers of items in two multiple-choice subtests. This made it difficult to make a fair comparison across the two sections and the levels of recognition and recall in this study. In designing future studies, careful attention should be paid to this aspect of test construction.

Furthermore, in this study, score reliability and criterion-related validity information were used as a central piece of evidence. Other types of validity evidence need to be sought on an ongoing basis. Continued research in this direction will surely enlighten us about the best ways to assess the EFL learners' knowledge of VPCs at all levels of the recognition-recall-production continuum.

References

- Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications*. (pp. 101-122). Oxford, England: Oxford University Press.
- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge, England: Cambridge University Press.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, England: Oxford University Press.
- Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice*. Oxford, England: Oxford University Press.
- Bahns, J. & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, 21, 101-114.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.
- Benson, M., Benson, E., & Ilson, R. (1997). *The BBI dictionary of English word combinations*. Amsterdam: John Benjamins Publishing Company.
- Bogaards, P. (2000). Testing L2 vocabulary knowledge at a high level: The case of the Euralex French Tests. *Applied Linguistics*, 21(4), 490-516.
- Bonk, W. J. (2001). Predicting paper-and-pencil TOEFL scores from KEPT data. *Kanda University of International Studies Research Institute of Language Studies and Language Education Working Papers*, 12, 65-86.
- Brown, J. D. (1995). *The elements of language curriculum: A systematic approach to program development*. New York: Heinle & Heinle Publishers.

- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671-684.
- Dagut, M. & Laufer, B. (1985). Avoidance of phrasal verbs: A case for contrastive analysis. *Studies in Second Language Acquisition*, 7, 73-79.
- Ellis, N. C. (2001). Memory for language. In P. Robinson (Eds.), *Cognition and second language instruction* (pp. 33 – 68). Cambridge: Cambridge University Press.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A.P. Cowie (Ed.), *Phraseology: Theory, analysis and applications*. (pp. 145-160). Oxford, England: Oxford University Press.
- Han, S. M. (2009). *Assessing Korean EFL learners' collocational competence: Item formats and receptive-productive dimension of vocabulary knowledge*. Unpublished master's thesis. Seoul National University, Seoul.
- Henricksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21, 303-217.
- Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, 19(1), 24 – 44.
- Lewis, M. (1993). *The lexical approach*. London: Language Teaching Publications.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Philadelphia: John Benjamins Publishing Company.
- Oxford University Press. (2000). *Oxford Collocations Dictionary for Students of English*. New York: Oxford University Press.
- Oxford University Press. (2004). *Really learn 100 phrasal verbs*. Oxford, England: Oxford University Press.
- Oxford University Press. (2007). *Really learn 100 more phrasal Verbs: Learn 100 frequent and useful phrasal verbs in English in six easy steps*. Oxford, England: Oxford University Press.
- Qian, D. D. & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing*, 21(1), 28-52.
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10(3), 355-371.
- Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. Kunnan (Ed.), *Validation in language assessment* (pp. 41-60). Mahwah, NJ:

Lawrence Erlbaum.

- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Runcie, M. (2002). *Oxford collocations dictionary for students of English*. Oxford, England: Oxford University Press.
- Schachter, J. (1974). An error in error analysis. *Language Learning*, 24, 205-14.
- Schmitt, N. (2004). *Formulaic sequences: Acquisition, processing, and use*. Amsterdam: John Benjamins Publishing Company.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford, England: Oxford University Press.
- Siyanova, A., & Schmitt, N. (2007). Native and nonnative use of multi-word versus one-word verbs. *IRAL*, 45, 119 – 139.
- Spearman, C. (1904). The proof and measurement of the association between two things. *American Journal of Psychology*, 15, 72 – 101.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Wesche, M. & Paribakht, T.S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *The Canadian Modern Language Review*, 53(1), 13-40.
- Worthen, B. R. (1993). Critical issues that will determine the future of alternative assessment. *Phi Delta Kappan*, 74, 444 – 454.
- Wray, A. (2000). Holistic utterances in protolanguage: The link from primates to humans. In C. Knight, M. Studdert-Kennedy, & J. Hurford (Eds.), *The evolutionary emergence of language*. (pp. 285-302). Cambridge: Cambridge University Press.

Hyojin Jeon

202-ho, 1690-59, Haengun-dong, Gwanak-gu,

Seoul, Korea 151-812

Phone: 82-10-6252-7577

Email: tohyojin7@naver.com

Received on 14 July, 2011

Revised version received on 31 August, 2011

Accepted on 31 August, 2011