

A Diagnostic Collocation Test for Second Language Learners

Alexandra H. Michaud & Siqun Dan

Northern Arizona University

Abstract

The following paper presents the test we designed to assess vocabulary knowledge of nine students enrolled in the Program of Intensive English at the Northern Arizona University. From the many dimensions involved in knowing a word, we chose to focus around the receptive collocation knowledge. We designed a diagnostic test made of multiple choice questions that would give insights to students and teachers about how comfortable students are with collocations of target words from their textbooks. After the administration, we ran an item analysis, we analyzed the descriptive statistics, and we addressed the issues of validity and reliability of the test.

Our results showed us that the test was not valid nor reliable. Some explanations are provided in the discussion about why the test didn't function as expected and about ways to improve the current version of the test.

A Diagnostic Collocation Test for Second Language Learners

Background

Vocabulary knowledge is the basic and indispensable element of language learning: “Without vocabulary nothing can be conveyed” (Wilkins, 1972, p. 111). Many researchers and teachers realized the significance of teaching vocabulary, but there is still a misunderstanding of how much learners need to know about vocabulary. Learning vocabulary is not just remembering the word’s meaning/definition; it also requires to interpret multi-layers of word knowledge which includes meaning, collocations, grammatical features, word parts and register/appropriate forms (Zimmerman, 2009). Researchers observed that second language learners, especially advanced learners, often face the problem of how to combine words together (Bahns & Eldaw, 1993). For example, some students cannot figure out whether they should use *spacious risk*, *taller risk* or *greater risk* in a paper or a conversation. This kind of knowledge, the words or types of words that occur with the keyword, is defined as collocation (Gyllstad, 2007). Collocation plays such an important role in words knowledge. However, in terms of students’ textbooks and teaching curriculum in Program of Intensive English (PIE), they are rarely explicitly taught by teachers. In order to explore the value of teaching collocation in the class, we designed a diagnostic collocation test for the L2 learners in PIE to assess their knowledge of collocation and the needs of teaching collocation in the class. With this test, students can at least have an idea of the amount of work they should put into practicing collocations. For teachers, they can decide how much time they should spend teaching collocation.

Research Questions

This paper is guided by three research questions: 1) Is the test valid? 2) Is the test reliable? 3) Does the test verify the necessity of teaching collocation?

Methods

Participants

The participants were nine students from the Program of Intensive English (PIE) at the Northern Arizona University (NAU). They were all either Chinese speakers or Arabic speakers. They had been enrolled in the level 3 (lowest level) of the PIE program since September 2016. The administration of the test took about 15 minutes, during which we answered students' questions about the test.

Instrument

The test was designed to measure the receptive collocation knowledge of the keywords in the textbooks. The construct is the collocation knowledge, which is the knowledge of the words or types of words that occur with the keyword. The subconstructs for this test include the ability to recognize the written form of the words and the knowledge of the words meaning, both for the target word and the distractors. This test is a diagnostic test that can be used as a reference for future collocation teaching. As students never learnt words' collocation explicitly before in the classroom and we want to testify how much they know, the test is a criterion-referenced test. 60% is the cut-point for pass- fail. All the vocabulary in the test is from the chapters covered by students' curriculum. (See the Appendix for a copy of the test.) With the help of Lex-tutor

frequency analysis and Antidote software, the correct option and distractor options were carefully chosen. Because all the words are picked according to the textbook and the Academic Word List, we expect a normal distribution of the scores.

Results

Item Analysis

The item analysis has proven to be a very good exercise for our test. We got a lot of information from the Item Facility (IF), Item Discrimination (ID) and B-index (Table 1). The IF shows that three items were easy the whole group and that three of them were quite difficult, one of them having an IF of 0. Looking at the ID, we observe that only five of the items (33.33%) discriminated correctly between the high and the low proficiency students (items 2, 5, 7, 8 & 14).

Table 1

Item Analysis

Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	TOTAL
8	1	0	1	1	1	1	1	1	1	0	0	0	0	1	0	9
4	0	1	1	0	1	1	1	1	1	0	0	0	0	1	0	8
2	0	0	1	0	0	1	1	1	0	0	0	0	1	1	1	7
6	0	1	0	0	0	0	1	1	1	1	1	0	0	1	0	7
9	0	0	1	1	0	1	0	0	1	1	0	0	0	1	0	6
1	0	0	0	1	0	0	0	0	1	1	0	0	1	1	0	5
3	0	0	0	1	0	0	1	0	1	0	1	0	0	1	0	5
5	0	0	1	0	0	0	1	1	1	1	0	0	0	0	0	5
7	0	0	0	1	0	1	0	0	1	1	1	0	0	0	0	5
IF	0,11	0,22	0,56	0,56	0,22	0,56	0,67	0,56	0,89	0,56	0,33	0,00	0,22	0,78	0,11	
IF upp	0,3	0,5	0,75	0,25	0,5	0,75	1	1	0,8	0,25	0,25	0	0,25	1	0,25	
IF low	0	0	0,4	0,8	0	0,4	0,4	0,2	1	0,8	0,4	0	0,2	0,6	0	
1 - IF	0,89	0,78	0,44	0,44	0,78	0,44	0,33	0,44	0,11	0,44	0,67	1,00	0,78	0,22	0,89	
IV	0,10	0,17	0,25	0,25	0,17	0,25	0,22	0,25	0,10	0,25	0,22	0,00	0,17	0,17	0,10	
ID	0,3	0,5	0,35	-0,55	0,5	0,35	0,6	0,8	-0,3	-0,6	-0,2	0	0,05	0,4	0,25	
IF pass	1	0	1	1	1	1	1	1	1	0	0	0	0	1	0	
IF fail	0	0,25	0,5	0,5	0,13	0,5	0,6	0,5	0,9	0,63	0,38	0	0,25	0,75	0,13	
B-index	1	-0,25	0,5	0,5	0,87	0,5	0,4	0,5	0,1	-0,6	-0,4	0	-0,3	0,25	-0,1	

High Average Low

The ID also shows that low proficiency students performed better on four of the items (4, 9, 10 &

11), and two items were achieved in more or less the same proportion by high and low proficiency students (12 & 13). The B-index tells us that seven items (46.67%) discriminated efficiently between students who passed the test and students who failed it (items 1, 3, 4, 5, 6, 7 & 8). However, it also indicates us that five of the items (33.33%) were better achieved by the students who failed the test (items 2, 10, 11, 13 & 15).

Descriptive Statistics

Looking at the descriptive statistics, we get a good idea of how the scores are distributed. With a mean of 6.33, a mode of 5.00 and a median of 6.00 out of 15, we can observe that the scores are pretty low. The mode is the lowest value of the distribution, which tells us that the distribution is positively skewed. That means that most of the scores of the distribution are low. The range is also pretty small (4) and the standard deviation is only 1.5, which means that the scores are concentrated towards the mean. The variance also shows us that there is not great variation within the distribution.

One of the most salient information that Figure 1 gives us is that the maximum score is 9, which also represents the cut-point. Figure 1 also shows us that the test taker that got the highest score is also the only one who actually passed the test.

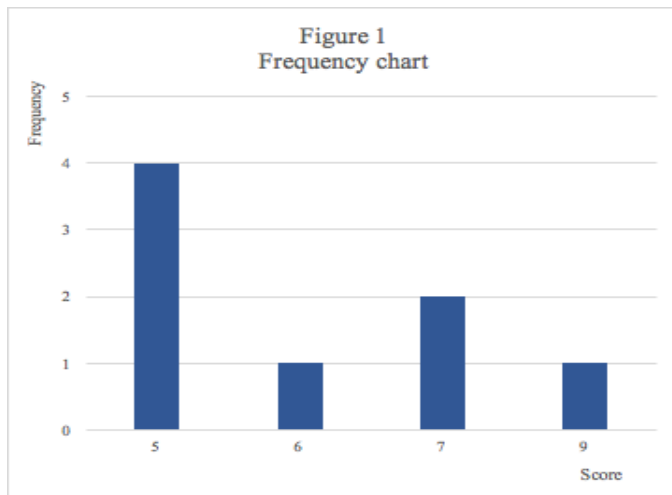


Figure 1. Frequency of scores on the collocation test.

To assess the reliability of our collocation test, we chose the KR-20 statistical test over the KR-21 because our items are obviously not the same level of difficulty thus we wanted the item facility (IF) to be taken into account. Our result for the KR-20 test was -0.20, which is even below the lowest possible result of 0. From this result we can draw the conclusion that the internal reliability or consistency was non-existent. Of all the different factors that could have influenced the reliability, our hypothesis are oriented around the number of test items and the number of test takers. If we would have had a larger sample, our KR-20 would probably have been a little higher.

Discussion

Unfortunately, the item analysis combined with the descriptive statistics and the reliability test have shown us that our collocation test is clearly not valid, at least for this sample of test takers. We were not able to design a test to assess the receptive knowledge of collocation. Most of the items didn't function as expected: they didn't distinguish between high and low proficiency students. However, even if the test itself didn't achieve its purposes, the main goal of this

experiment was to assess the validity of the test, and for this purpose we got a pretty clear answer that it didn't. The KR-20 is negative, which infers the low reliability. More test takers would probably have increased the reliability level. Since the B-index of five items is negative, these items are not valid.

We don't necessarily think that the items themselves were problematic. The distractors were carefully chosen: all the words (target words, collocate, distractors) were from the same frequency band (1K to 3K), a native speaker reviewed the test to make sure the distractors were not possible answers, and the target words were technically known by test takers since they come from the chapters that students studied during the semester. However, we could have added the definitions of the target words and distractors to make sure that students knew at least the meaning. There is a possibility that some test takers didn't even know the meaning of the target words and that they probably guessed any answer. Since knowing the meaning and knowing the use are two different dimensions of knowing a word (Nation, 2001), we don't think that it would have biased the answers. It would have been totally feasible to design a web quiz where test takers can access the words' definition when needed.

Relevance to PIE and Second Language Learning

The goal of this project was to assess the receptive collocation knowledge of L2 English and discuss the need of collocation teaching. Knowing that this test was meant to be a diagnostic test, it can still give some valuable information about how students from the level three of the Program of Intensive English at the Northern Arizona University are comfortable with collocations of

vocabulary they technically know. As the results indicate, the low pass rate of the test tells us about the limited collocation knowledge of learners. In other words, teachers need to raise attention of collocation and at least let students notice their limitation and learn the collocation inside or outside the classroom. Looking back at the research questions, we conclude that the test was neither valid nor reliable, and that it could give valuable insights for collocation practice but probably for students of a higher level.

References

- Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations. *System*, 21, 101-114.
- Boyd, C. (2009). *Word knowledge: A vocabulary teacher's handbook*. New York, NY: Oxford University Press.
- Gyllstad, H. (2007). *Testing English collocations: Developing receptive tests for use with advanced Swedish learners*. Lund University.
- Miller, M. D., Linn, R., & Gronlund, N. (2013). *Measurement and assessment in teaching (11th ed.)*. New York, NY: Pearson.
- Wilkins, D. 1972. *Linguistics and language teaching*. London: Edward Arnold. Zimmerman, C.

Appendix

Vocabulary test

Vocabulary test on collocations, based on Program of Intensive English (PIE) corpora

Choose the answer that represents the best association between the capitalized word and the other words.

Only one possible answer per question

1- Verb + COMPETITION

Mark only one oval.

- Face competition Make competition
- Do competition
- Approach competition

2- Verb + CONTROL

Mark only one oval.

- Catch control
- Do control
- Take control
- Make control

3- Adjective + IMAGINE

Mark only one oval.

- Hard to imagine
- Bad to imagine
- Heavy to imagine
- Confusing to imagine

4- Adverb + INVOLVED

Mark only one oval.

- Bigly involved
- Deeply involved
- Much involved
- A lot involved

5- Verb + CHALLENGE

Mark only one oval.

- Face a challenge
- Be in a challenge
- Play a challenge
- Do a challenge

6- EVIDENCE + verb

Mark only one oval.

- Evidence thinks
- Evidence does
- Evidence wants
- Evidence suggests

7- Verb + MODEL

Mark only one oval.

- To do a model To
- build a model To
- grow a model To
- rise a model

8- SPECIFIC + Noun

Mark only one oval.

- Specific bother
- Specific particulars
- Specific abandon
- Specific details

9- Adjective + DEMAND

Mark only one oval.

- Admired demand
- Favoured demand
- Popular demand
- Well demand

10- INCREASE + adverb

Mark only one oval.

- Increase fast
- Increase rapidly
- Increase ma
- increase much

11- Adjective + INCOME

Mark only one oval.

- Annual income
- Fake income
- Timed income
- Year-long income

12- Adjective + RISK

Mark only one oval.

- Spacious risk
- Voluminous risk
- Taller risk
- Greater risk

13- Adverb + EFFICIENT

Mark only one oval.

- Widely efficient
- Highly efficient
- Best efficient
- Well efficient

14- Verb + PREDICTION

Mark only one oval.

- Take a prediction
- Have a prediction
- Make a prediction
- Do a prediction

15- Adjective + DISEASE

Mark only one oval.

- Deep disease
- Severe disease
- Profound disease
- Great disease