

Title	The JWL 8000 and the power of distraction
Sub Title	JACET8000と語彙テスト 錯乱肢
Author	中村, 優治(Nakamura, Yuji) Murray, Adam
Publisher	慶應義塾大学日吉紀要刊行委員会
Publication year	2012
Jtitle	慶應義塾大学日吉紀要. 言語・文化・コミュニケーション (Language, culture and communication). No.44 (2012. ),p.73- 82
JaLC DOI	
Abstract	
Notes	
Genre	Departmental Bulletin Paper
URL	<a href="https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN10032394-20121231-0073">https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN10032394-20121231-0073</a>

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

# The JWL 8000 and the Power of Distraction

Yuji Nakamura  
Adam Murray

## Introduction

It can be argued that vocabulary knowledge is essential to measuring proficient reading levels in both first and second languages. While there has been extensive discussion about the importance of teaching vocabulary in academic settings, testing vocabulary itself has received less attention. Currently, the validity of vocabulary testing as a distinct construct is a matter of debate, whereas the assessment of reading and reading comprehension has consistently been linked to some form of vocabulary measurement (Ekbatani, 2011).

Milton (2009) argues that when vocabulary is being tested, the essence of the words should be apparent to the examinee. When constructing a test that seeks to measure vocabulary knowledge, it is essential for the test to explicitly define what is being measured. Furthermore, in designing the exam, a thorough understanding of how vocabulary knowledge can be effectively measured and what constitutes a good test should be investigated.

Read (2012) also claims that it is necessary to define the scope of vocabulary testing. Vocabulary is obviously an indispensable component of language use, to the extent that almost any kind of language test is in a sense measuring the test-takers' vocabulary ability, even if this ability is not an explicit focus of the scoring or rating.

Read (2000) presents two opposing viewpoints on the role of vocabulary in language assessment. One view suggests that it is entirely reasonable to create tests that measure vocabulary meaning and the correct usage of a set of vocabulary items. The other view is that vocabulary can only be assessed in the context of a language-use task, where vocabulary interacts in a natural way with other elements of language

knowledge. We need to consider both of these viewpoints when constructing academic tests.

## **Purpose of the Study/Research Questions**

The purpose of the present study is two-fold: 1) to discuss important issues in second language vocabulary assessment in academic settings and 2) to analyze the vocabulary section of a university placement test.

In general, there are two approaches to define 'difficulty' or how difficult it is for EFL learners to acquire vocabulary. One method is based on the 'word level', which is defined as how necessary a word for reading. The other is based on the testing approach, in which vocabulary difficulty is explained by test results. In this approach, reliability and validity measurements are necessary. To do this, a large-scale test was administered to construct an item difficulty scale for the words. We estimated a standardized item difficulty parameter for each item to compare the level of difficulty in a common scale through placement tests (PTs) and confirmation tests (CTs) from 2006 to 2010. Additionally, we counted the frequency of each 'word level' for each PT and CT test form. Thus, we can describe the relationship between these two scales using PTs and CTs.

We set three research questions (RQ) for this study, the first being: what is the relationship between the word level of distractors and their attractiveness? Are low-level words less attractive distractors? Also, are higher-level words more attractive distractors? These will be the effective questions, because it is convenient for the test maker to describe the general behavior of the low-level English learner. To answer RQ1, we adopted the JACET8000 word level list (JWL), which is also used later on in this study.

The second RQ is: Are Off-list words the most enticing distractors? In the JWL, there are nine levels of difficulty, from 1 to 8, and a 'plus' level that is set under level 1. However, there are quite a few words that do not appear in the JWL. Generally speaking, a word list will contain the words that are important and are frequently used in materials used by English learners. Each word's rank is determined by its importance. Then, it is natural to conclude that the Off-list words are less important than those of the JWL. This RQ is linked to RQ1. If the answer to RQ1 is 'higher-level words are more attractive distractors', then the answer to RQ2 must also be true.

Because a scale based on word level can be considered to reflect English reading skills, our placement tests also measure English reading skills. However, the construct validity for the vocabulary section has not been fully examined yet. In this

study, our placement tests do not measure vocabulary skill directly, but assess vocabulary skill by reading English documents that can be seen in college or university classrooms. If the word level approach does not match the test scale, there might be a need for different scales to measure different factors of vocabulary skill. On the other hand, we assessed vocabulary skill by having the test takers read passages in English. As a result, a readability score, such as the Gunning-Fog index, is a good predictor for explaining item difficulty. For this reason, we set RQ3: Are readability (Gunning-Fog index score) and word level (JACET 8000) good indicators of the difficulty of vocabulary test items? This RQ is mainly set for the purpose of examining the construct validity for placement tests, and partly set for the purpose of exploring how difficulty may be described.

## Method

### *Subjects*

Keio University Faculty of Letters began administering an in-house placement exam for incoming freshmen and new sophomore students in the spring semester of 2006. Students are given placement tests twice a year, once at the beginning of the academic year and again in the fall. For this analysis, the results of 10 exams (approximately 800 test-takers each) were examined.

### *Materials/Instruments*

The placement test is a 60-minute examination which consists of 50 questions in four sections: grammar, vocabulary, gap-fill (cloze), and reading. For this study, only the vocabulary section was examined. The vocabulary section consists of 10 multiple choice questions with four options. The contents of 10 placement tests, 100 test items in total, were examined.

### *Analysis Procedures*

First, we made a list of the words that appeared in the items in all 10 placement test forms. Each item has four options, so the list consisted of 400 words. Next, we checked the JACET8000 level of each word. (Analysis 1).

Second, using item response theory (IRT) we constructed an item difficulty scale using WINSTEPS. We considered this scale to be a measurement of vocabulary skill in reading English documents. In the IRT analysis, 92 test items, were examined. The total number of test items was not 100 because each test form contains common items (anchor items) to apply common-item nonequivalent groups design using IRT analysis. However, because the results were different among test forms for the same anchor item, an analysis of attractiveness of options was conducted for all 100 items.

In analysis 2, we examined the distribution of item difficulty by examining the JACET8000 level of the four options for each test item. This analysis provided the answer to RQ2. Note that not all 400 words were used in this analysis. Because there were multiple appearances of words among the 100 items, we excluded these recurring words.

Third, for items which fewer than 40 percent of the students answered correctly, we calculated the percentage of each option selected by the test takers. The JACET8000 level for each option was also considered (Analysis 3). This provided us the answer to RQ1.

Finally, for all 100 items we calculated the Gunning-Fog index score for readability. Also, a correlation analysis was conducted to examine the relationships between the Gunning-Fog index score and item difficulty, item discrimination, and JACET8000 level to answer RQ3.

## Results and Discussion

Table 1 shows that a total of 400 words (including overlapping words) were used in this analysis. Of the 9 different word levels of the JACET 8000, Level 1 and Plus words were the most frequently used (27%), followed by Off-list words (18%), then by Level 2 (12%). Levels 6, 7, and 8 were the least frequently used.

Table 1  
*Distribution of JWL Words*

	A	B	C	D	Total	Proportion	%
Level 1	33	28	21	17	99	0.2475	25
Level 2	12	9	10	19	50	0.125	12
Level 3	6	5	7	15	33	0.0825	8
Level 4	5	17	12	9	43	0.1075	10
Level 5	15	8	12	5	40	0.1	10
Level 6	5	6	2	5	18	0.045	5
Level 7	2	6	11	2	21	0.0525	5
Level 8	4	5	3	6	18	0.045	5
Plus (=level 1)	1	1	3	1	6	0.015	2
Off-list (level 9)	17	15	19	21	72	0.18	18
	100	100	100	100	400		100

*Note.* Includes overlapping words.

Table 2 and Figure 1 indicate that words in each JWL have various levels of difficulty. In Level 1, there are 11 items which range in difficulty from 32.7 to 56.9. However, at Level 8, the three items range in difficulty from 34.1 to 53.7. This means

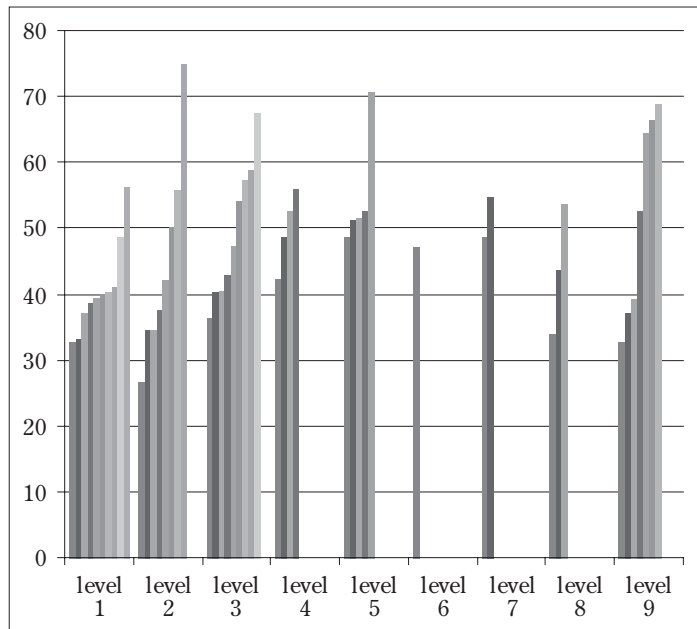
that it is not necessarily true that higher-level items are more difficult than lower-level items. Each of the 9 levels contains both easy and difficult items. In other words, the higher-level words could be easier test items. In fact, one of the easiest items was a Level 9 word (32.8). Conversely, there were some lower-level items which were very difficult, with the most difficult item being a Level 2 word (75.1).

Table 2  
Distribution of Items by Difficulty in Each JWL

Difficulty	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Level 8	Level 9
Easy	32.7	26.7	36.4	42.5	48.8	47.1	48.8	34.1	32.8
	33.4	34.5	40.5	48.7	51.4		54.8	43.8	37.3
	37.3	34.7	40.7	52.7	51.8			53.7	39.3
	38.9	37.8	43.0	56.2	52.8				52.6
	39.7	42.1	47.6		70.7				64.6
	40.2	50.4	54.2						66.5
	40.5	56.0	57.4						69.0
	41.1	75.1	59.0						
	48.9		67.6						
	56.5								
Difficult	56.9								

Note. Excludes overlapping words.

Figure 1  
Distribution of Items by Difficulty in Each JWL



Note. Excludes overlapping words.

Table 3 shows the attractiveness of options by JWL. Low-level words (Level 1 and Plus-level) can be very attractive distractors. For PT1(24), 53% of the test-takers selected “ran”, a Plus-level distractor. Similarly, for PT1(23), 60% of the test-takers chose “turn”, a Level 1 distractor. PT6(16) and PT8(21, 22) also contain very attractive Level 1 distractors, such as “way” and “trouble”. High-level words (Level 9) can also be very attractive distractors. PT5(21,22) and PT6(17) show that Level 9 words such as “submissively” and “originality” are also very attractive. Also, mid-level words (Level 5) can be attractive distractors. Examples of this are PT2(18), PT4(22), PT5(25), PT6(25), PT8(19) and PT9(18) which words such as “applicable”, “postponed” and “alternatively”. In short, both high-level (including Off-list words) and low-level (including Plus) words can be either more attractive or less attractive distractors.

Table 3  
*Attractiveness of Options by JWL*

Test	Item #	Proportion of Selection				JWL			
		A	B	C	D	A	B	C	D
PT1	Item 19	28	<u>26</u>	23	22	1	1	9	5
PT1	Item 23	60	11	3	<u>25</u>	1	1	1	3
PT1	Item 24	8	30	53	8	1	9	plus	1
PT2	Item 18	24	<u>21</u>	35	21	1	7	4	1
PT2	Item 23	7	<u>23</u>	64	6	1	9	plus	1
PT3	Item 22	7	<u>35</u>	48	9	1	9	plus	1
PT4	Item 19	24	46	<u>21</u>	8	9	2	9	2
PT4	Item 22	<u>20</u>	59	10	10	5	5	7	7
PT5	Item 21	<u>29</u>	4	45	21	9	6	9	6
PT5	Item 22	37	9	<u>15</u>	38	9	2	2	5
PT5	Item 25	<u>20</u>	57	14	8	5	5	7	4
PT6	Item 16	<u>36</u>	14	6	43	1	2	3	1
PT6	Item 17	46	17	<u>7</u>	29	9	4	6	8
PT6	Item 25	10	39	27	<u>24</u>	9	5	5	5
PT7	Item 25	<u>10</u>	22	15	53	1	plus	1	plus
PT8	Item 19	17	<u>30</u>	39	13	3	7	5	3
PT8	Item 21	<u>17</u>	52	12	18	8	1	5	8
PT8	Item 22	11	52	14	<u>23</u>	1	1	1	1
PT9	Item 18	21	<u>33</u>	33	13	3	7	5	3
PT10	Item 24	50	14	<u>28</u>	8	4	5	4	5

*Note.* Underlined selections are the correct answers.

Table 4 provides information about the attributes of 30 test items. Using PC as a relative measure of difficulty, these test items were divided into three categories: easy, medium, and difficult. Easy items have a PC greater than 80, medium items have a PC between 40 and 80, and difficult items are those with a PC below 40. Some

difficult items have low-level words but have high Gunning-Fog index scores indicating difficult item prompts. However, some difficult words are both low-level and have low Gunning-Fog index scores, but are categorized as items relying on collocation knowledge (Category A) or requiring vocabulary depth (Category C). There are easy items where word level is high and the Gunning-Fog score is high, but are categorized as items requiring vocabulary breadth (Category D). (For examples of each of the four categories, see Appendix.) In addition, there are some easy items with high-level words with low Gunning-Fog scores. There are some difficult items with low-level words and low Gunning-Fog scores but require vocabulary depth knowledge. When poor students find the higher-level items easy, the prompt was easy to read. When good students find the low-level items difficult, the prompt was difficult to read.

Table 4  
*Attributes of Easy, Medium, and Difficult Test Items*

JWL	Fog	PC	PBs	Category
1	2.4	97	20	d
9	11.31	93	42	b,c
2	6.4	90	43	c
1	6	90	27	c,d
1	11.47	90	28	c,d
1	4	87	21	b,c
1	8.13	86	31	d
1	8.13	86	32	d
8	3	86	30	a,c
2	2.2	86	34	c,d
1	5.6	61	25	b
7	9.07	61	34	d
7	9.07	61	29	d
5	5.2	60	34	c,d
5	5.2	59	34	c,d
5	5.2	58	35	c,d
2	8.04	57	28	a,d
5	13.2	57	28	a,c
5	16	55	45	c
8	8.46	54	24	b,c
3	8.67	25	17	b,c
5	8.04	24	24	a,c
9	11.35	23	29	a,b
7	9.15	21	6	b,c
9	14.43	21	16	a,b
5	8.67	20	23	d
5	8.67	20	10	d
8	2.8	17	15	a,c
2	11.35	15	9	c,d
plus	3.6	10	15	c,d

*Note.* Categories: A-collocation, B-grammar, C-vocabulary depth, D-vocabulary breadth.



Table 5 shows that there is a significant correlation between the JWL level and the item difficulty (PC). In other words, the higher the level is, the more difficult the item. However, there is very little correlation between readability (Fog) and difficulty (PC). There is a weak correlation between the JWL level and item discrimination (PBs).

Table 5  
*Correlations Between JWL, Readability, Difficulty, and Discrimination*

		Level	Fog	PC	PBs
Level	Pearson Correlation	1	.198*	-.271**	-0.028
	Sig. (2-tailed)		0.049	0.006	0.782
	N	100	100	100	100
Fog	Pearson Correlation	.198*	1	-0.181	0.088
	Sig. (2-tailed)	0.049		0.071	0.385
	N	100	100	100	100
PC	Pearson Correlation	-.271**	-0.181	1	.594**
	Sig. (2-tailed)	0.006	0.071		0
	N	100	100	100	100
PBs	Pearson Correlation	-0.028	0.088	.594**	1
	Sig. (2-tailed)	0.782	0.385	0	
	N	100	100	100	100

Note. \*p<.05. \*\*<.01.

## Answers to Research Questions

RQ1: Are low-level words (distractors) less attractive distractors?

Some low-level words (Levels 1, 2 and Plus) were attractive distractors, and thus the research question was not strongly supported.

RQ2: Are higher-level words (distractors) more attractive distractors?

Some higher-level words (Levels 7, 8) were attractive, so the research question was not strongly supported.

RQ3: Are Off-list words the most enticing distractors?

Level 9 (Off-list) words were attractive distractors in very few cases, so this research question was not strongly supported.

RQ4: Are readability (Gunning-Fog index score) and word level (JACET 8000) good indicators of the difficulty of vocabulary test items?

There is very little correlation between readability and difficulty, or between JWL and difficulty, thus readability and word level cannot be considered good indicators of difficulty.

## Conclusions and Implications

From this study we can conclude that not all factors of vocabulary can be measured by only one type of test. When preparing vocabulary test items, word level, readability, and the contents of the item should be taken into consideration. Also, attention must be paid to the clear meaning of categorization of vocabulary terms such as vocabulary size, breadth, depth, collocation, phrases, and idioms.

Read (2012) states that one trend in vocabulary studies is the increasing interest in collocations and other kinds of multi-word lexical items. He points out that some types of multi-word lexical items can be assessed in the same way as single words, but others will require contextualized test formats. Further study is needed to examine the construct validity of vocabulary testing for EFL.

### Acknowledgement

This work was supported by the Grant-in-Aid for Scientific Research (KAKENHI) (C) (24520646).

### References

- Assessment Systems Corporation. Xcalibre 3 [computer software] <http://www.assess.com>
- Douglas, D. (2010). *Understanding language testing*. London, UK: Hodder Education.
- Ekbatani, G. (2011). *Measurement and evaluation in post-secondary ESL*. New York: Routledge.
- Fulcher, G. (2010). *Practical language testing*. London, UK: Hodder Education.
- Gunning fog index. (n.d.). In *Wikipedia*. Retrieved April 21, 2011 from [http://en.wikipedia.org/wiki/Gunning\\_fog\\_index](http://en.wikipedia.org/wiki/Gunning_fog_index)
- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge, UK: Cambridge University Press.
- JACET (The Japan Association of College English Teachers). (2003). *The JACET List of 8000 Basic Words*. Tokyo: JACET.
- Linacre, L. (2005). WINSTEPS (Version 3.55) [Computer software]. <http://www.winsteps.com/winsteps.htm>.
- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol, UK: Multilingual Matters.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.
- . (2005). Teaching and learning vocabulary. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 581–595). Mahwah, NJ: Lawrence Erlbaum Associates.
- . (2008). Teaching vocabulary: strategies and techniques. Boston: Heinle.
- O'Sullivan, B. (ed.). (2011). *Language Testing: Theories and Practice*. Oxford, UK: Palgrave Macmillan.

- Read, J. (2000). *Assessing vocabulary*. Cambridge, UK: Cambridge University Press.
- . (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics*, 24, 146-161.
- . (2012). Piloting vocabulary tests. In G. Fulcher & F. Davidson (Eds.), *The Routledge Handbook of Language Testing* (pp.307-320). UK: Routledge.
- Schmitt, N. (2010). *Researching vocabulary: a vocabulary research manual*. Oxford, UK: Palgrave Macmillan.

### *Appendix*

#### Sample Test Items for Each Category

#### Collocation (A)

They had to ( ) very early to avoid the traffic jam.

- ① set off                      ② cut off                      ③ call off                      ④ put off

Our teacher makes it a ( ) to start his class on time.

- ① diary                      ② rule                      ③ date                      ④ lesson

#### *Grammar (B)*

He can give you one or two ( ) of advice about your essay .

- ① slices                      ② lumps                      ③ bites                      ④ pieces

#### *Vocabulary Depth (C)*

Don't take his words ( ). You had better read between the lines.

- ① accurately                      ② faithfully                      ③ literally                      ④ honestly

I cannot concentrate on my studies. I'm ( ) with job hunting.

- ① possessed                      ② taken away                      ③ devoured                      ④ preoccupied

#### *Vocabulary Breadth (D)*

If you feel any pain, ( ) these pills. They work very well.

- ① pick                      ② drink                      ③ take                      ④ taste

What's the ( ) of this news? Is it really reliable?

- ① supply                      ② demand                      ③ source                      ④ portion