

---

---

## Research Digest

# Practicality and Validity in Three Achievement Tests of Vocabulary

George M. Harrison

*College of International Relations*

*Nihon University, Mishima*

Teachers wanting to bring more cohesion into a course, or who want to be informed of how well their students have learned vocabulary taught in class, often write their own short-term achievement tests. Because these tests are often used only one time, and because they should be fair in terms of the language construct being tested, a teacher is often faced with the task of creating test items which are not only easy to write and score, but which approximate as best as possible vocabulary knowledge studied in class (Nation, 2001). Many teachers developing classroom tests, myself included, often go through this process intuitively, without much systematic attention to the effort that goes into developing, administering, and scoring a test, or to how well the test does its job. The purpose of this study, therefore, is to find out, given a choice of three vocabulary test formats, which is most usable in terms of practicality and validity.

### Three Test Designs

*Student-centered test.* Nation (2001, p. 375) presents a test which deals with a variety of word knowledge aspects. With this weekly test,

the teacher chooses 10 words for the class which will be on the test; in addition, each student chooses their own 10 words ( $k_{\text{total}} = 20$ ), and gives them to the teacher who then directs the student to either write a sentence, write the word's collocates, its definition, or derivations of the word (same word family).

*Word association test.* Read (2000, 2004) presents a selective word association test he developed to measure students' depth of knowledge of items in Xue and Nation's (1984) University Word List (UWL). As with traditional productive word association tests, it included paradigmatic and syntagmatic relationships. Read had originally set out to measure depth of knowledge of items in the UWL and had established concurrent validity (correlating it with a definition matching test) and reliability (using Rasch modeling). He found it was difficult, however, to make the items consistent in how the associations related to the prompt and therefore limited his items to one word form, adjectives. He also found guessing to be a big factor with individual items because learners could identify connections among the item choices. In response, he developed a format where some items had two correct choices in each box, while others had one correct choice on one side and three on the other. Here are example test items with two correct answers in each box:

### Sudden

beautiful	quick	surprising	thirsty
-----------	-------	------------	---------

change	doctor	noise	school
--------	--------	-------	--------

### Common

complete	light	ordinary	shared
----------	-------	----------	--------

boundary	circle	name	party
----------	--------	------	-------

Read, 2000, p. 184

Read recognized that good guessing may still be responsible for much variance, and in looking at the items, we can see that if a learner did not know the meaning of *sudden*, she could look at the choices

and see that paradigmatic choices such as *beautiful* and *thirsty* are unlikely matches with some syntagmatic choices such as *doctor* or *school*. Developing distractors that avoid this may be difficult, and even more so if the choices are to be limited to the first 2,000 most frequent words.

*Sentence completion tasks.* Read (2000) also describes a blank-filling task as one type of classroom progress test that requires learners to draw from the immediate context to produce the correct word. In the blank, the initial letter is provided to reduce cognitive load. Here are two of Read's examples (p. 174):

- a) Many diamond mines are l\_\_\_\_\_ in South Africa.
- b) Modern jetliners fly at an a\_\_\_\_\_ of 35,000 feet.

These items appear to test students' ability to recall and produce a word (lemma) given a context (i.e., the conceptual and grammatical hints in the sentence). Since learners are expected to use vocabulary studied in class, Read justifies not using multiple initial letters when other responses are possible.

Comparing these three tests in just one area may not be a fair assessment of their use. For example, the student-centered progress test appears easiest to develop, but may be problematic for scoring and validity. Analyzing these three tests in terms of the practicality of developing, sitting, and scoring them and in terms of their perceived validity within the context of their use would be more revealing. I pose the following research questions:

How do the three tests rank against each other in terms of

- a) the resources required (for making, taking, and marking the test)? (RQ<sub>1</sub>)
  - b) how well they appear to test the knowledge studied in the textbook? (RQ<sub>2</sub>)
-

This is an effort, in other words, to determine which test is most usable for this context in terms of (a) practicality, and (b) validity.

## **Methodology**

To answer these questions, a sample of 12 of the original 34 vocabulary items taught were selected by taking every third one from the order in which they occurred in the course textbook (Harrison, 2007). Each of the three tests (see Appendix) contained these 12 items. The “student-centered” test was compromised to some extent because the choice of items was not the students’. Still, this was justifiable for practical reasons: (a) to be able to compare the three tests, and (b) the class in which I used the textbook ended 10 weeks prior to the study.

To help in choosing distractors for the word association test and the words in the sentence completion test, I consulted an electronic dictionary containing the Oxford Dictionary, Thesaurus, and Collocation databases. For the vocabulary not studied in the textbook and which were not loanwords in Japanese, I checked the General Service List (West, 1953) to see if the distractors I came up with were in the first 2,000 most frequent words. If not, I searched for a new distractor.

The word association test underwent the most adaptation from the original test design it was based on. In items with multi-word units (MWUs), three modifications were made: (a) nodes (sometimes more than one word) were identified and used as prompts, (b) choices included multi-word parts that made up the unit (sometimes with ellipses cueing which side of the choice the syntagmatic relationship lay), and (c) the number of correct choices per item varied (in an effort to hasten the test-writing and reflect the various meanings addressed in the textbook). Efforts were made to keep the choices to the same part of speech or grammatical unit, which worked fine with the paradigmatic associations, but was impractical for the syntagmatic choices.

The three tests were given to each participant in the following order:

(a) the sentence completion test, (b) the student-centered test, and (c) the word association test. This order was chosen to minimize test effect. After each test, I conducted a group interview, asking students how they arrived at their answers. At the end, I asked general impression questions about the three tests.

*Participants.* The tests were given to two male third-year Japanese university students from an advanced English course which ended 10 weeks prior to the study. They were selected because I had used my textbook in that class and they would more closely resemble students who had just studied the material than would students who had never encountered the text. Both students are motivated to learn English and are members of the campus English club. One student, Y (short for Yuki; pseudonyms used), had reviewed the textbook during the 10 week break.

## **Results and Analysis**

*Practicality.* Table 1 shows the results for the time and effort put into the test. In terms of time, the student-centered test looked to be superior, with the sentence completion test in second, and the word association test last. Because the tests were given to only two students, the time in marking each is extrapolated to 20 students (actual time times 10), a conservative estimate compared to many university English classes in similar contexts. Time to mark, therefore, is a very rough estimate of the time it would take a teacher to score the tests and should be considered carefully.

In terms of effort, the sentence completion test required the least effort, with the student-centered test next, and the word association test once again in last place.

Table 1. Time and Effort Resources

Test	Time (minutes)				Effort perceived (scale of 1 to 10, 1 = least effort)				Other resources (to make)
	make	take	mark <sup>1</sup>	total	make	take <sup>2</sup>	mark	Avg.	
Sentence Completion	65	15	15	95	4	4	2	3.33	electronic dictionary
Student- centered	10	17	50	77	1	5.5	10	5.50	
Word Association	135	10	20	165	9	2	8 <sup>3</sup>	6.33	electronic dictionary

<sup>1</sup>Results for two test candidates, but extrapolated to 20 candidates

<sup>2</sup>Based on students' judgments

<sup>3</sup>This could have been made easier if there were a systematic number of correct choices.

Since time and effort are different scales, I cannot give an overall quantitative ranking of the tests, making it difficult to provide a definitive answer to the first research question. Also, each of the three scales (making, taking, and marking) cannot be assumed to be equal in value, although that is how I treated them.

I can say, based on these data, that the word association test was the least practical for writing; it took a great deal of time and mental effort to develop distractors that were semantically or grammatically likely, but which were also within the 2,000 most frequent words. The student-centered test took very little time and effort to create, but was difficult and time-consuming to grade due to some ambiguous responses. The sentence completion test seemed the most balanced in terms of time and effort it took to create, sit, and score. In similar classes with close to 30 students, the sentence completion task test may be most practical; for classes less than 20 students, the student-centered test would be a better choice.

*Validity.* The first test, the sentence-completion test, measured productive recall of the word or MWU based on the context provided. Some of the student comments in response to the question, "What were you thinking when you chose your answer" are shown in Table 2.

Two things can be said based on the data and test results: (a) test takers engaged semantic and grammatical clues in context to help them arrive at answers, and (b) sentences that were similar to those in the textbook or which were true for students' lives were easier than sentences that do not fit these situations.

The student-centered test measured receptive recall (of the items), but some language production was needed to complete the test tasks. Student comments answering the questions, "Which were easy?" "Which were difficult?" and "Why?" can be found in Table 3.

Table 2. Interviewee Comments After Sentence-Completion Test

Item	Interviewee comments	Interpretation
complicated	H: "It was easy. We can see 'simple' and 'but' and 'relationship'."	Semantic and grammatical clues
	Y: "Yeah, we studied this in the class."	Textbook context was same and Ss remembered it
tease	H: "It's a negative word because of 'lonely'. At first I thought 'bully' or 'ignore,' but the first word is 't' so I thought 'tease'."	Semantic and grammatical clues
	Y: "The situation."	Textbook context was same and Ss remembered it
stress	T: "I thought why they do exercise is to relieve stress, and that's what I do."	Connected context to personal experience
		Semantic clue; possibly collocation
commute	H: "We studied this word, but I can't remember it."	
	Y: "To and from work, so that means the time to come to a place or back to a place. I want to live closer [to school], so from the sentence, it's true for me."	Textbook context was the same, but Ss couldn't recall form
	(Note: Y got partial credit for the synform "commit")	Connected context to personal experience
effort	Both Ss realized their answer is not correct. Note: The collocation in the test, "make more of an effort," is longer than the textbook's "make an effort."	The generative use may have made the item more difficult

Table 3. Interviewee Comments After Student-Centered Test

Item	Interviewee comments	Answer produced	Interpretation
complicated	H: "I know this word"	"I have so much works in a day that things are complicated" [sentence]	Ss know the word and the situation
	Y: "I'm a little confused how to use the word 'stuff.'"	-	lack of concrete meaning may have been difficult to remember
I hear you	H: "difficult"	-	S remembers an instance of use.
	Y: "I remember using this at the pub for your party"	"I understand you." [definition]	H didn't use the same sense that I was testing, but Y did.
take long	-	H: "I live close to this uni. It doesn't take long to come to here." Y: "It didn't take long for me to notice that my friends break the promise." [sentence]	-
	Y: "It was easy for me to make the sentence because it's a useful word and I use the word in daily life."	Y: "You should make an effort to be loved by your girlfriend." [sentence]	Y is drawing on his life experience.

Students said, after taking all three tests, that they liked this version the best. When asked to elaborate, Y said, "It's fun if I can show I know the meaning of the word," and H (Hiro) said, "Because we have to know the exact meaning of the word and it's useful for us in daily life." Several comments related to the fact that they felt they were given more of a chance to show what they knew about the word compared to the other two tests.



Students viewed the student-centered test as most similar to the way they are expected to use the language in real life. In terms of target language use domain, it was similar to several tasks in the textbook. Furthermore, the sentences and definitions produced by one student, Y, were similar to those in the book, but those of H, were mostly different. Since student Y reviewed the textbook over the 10-week break, but H did not, at least some degree of validity can be established. If the test were used in the middle of a course, as it should be, perhaps the responses of student H would have more closely resembled those of the textbook, and he would have got a better score.

This test clearly separated the meaning and use aspects of vocabulary (allowing me to control which to measure), but presented difficulties in grading. It was difficult, for example, to decide whether to accept answers that related to a different sense of the word than what I was looking for. In the middle of a course, the teacher may explicitly instruct students to use the sense that was studied, or test the underlying meaning. Another difficulty I had lay in the ambiguity in drawing the line between the grammaticality of the use of the item and that of the rest of the sentence. It seemed unfair to penalize grammar that was not related to the item, but it was sometimes necessary when it was in close proximity to the item.

The word association test measured receptive recognition in the two scales of meaning (paradigmatic use) and use (syntagmatic use). Student comments are listed in Table 4. This test appeared to successfully measure students' knowledge of the senses of the words in the meaning scale. H, for example, explained that he knew one sense of "get to" and "I hear," but not the one I was testing. Y, who had been reviewing the textbook, correctly answered these items. Since the textbook directly taught these new senses, the test achieved its goal.

Table 4. Interviewee Comments After Word Association Test

Item	Interviewee comments	Interpretation
gossip	H: "It's natural to associate 'gossip' with 'talk about' but I think sometimes it also may mean 'laugh'"	H knows the meaning of the word, but is associating what people do when they gossip (laugh).
	Y: "I wonder if there are two or more answers, even though I chose one for each. Sometimes I chose two."	Test familiarity; also need to systematize number of correct As
I hear get to	H: "Some of the words ['I hear,' and 'get to'] I couldn't understand the meaning."	H knows one sense of the word, but not that which is tested.

Word associations seemed to work relatively well even when some items confounded grammatical functions and collocations among their syntagmatic choices. However, for one item, *gossip*, one student was convinced that *laugh* should be a paradigmatic association. This relates to the non-sensitivity of the test; i.e., in an effort to find plausible choices, many of my distractors were similar in meaning.

## Discussion and Final Comments

Several limitations emerged in developing and using these three tests. First, this may not be a completely accurate estimation of the tests' usability in their intended context because the participants selected for the project were not, at the time of the study, in a class using this textbook. More students would also be more desirable for a valid analysis. Second, in future studies, the word association test should have a systematic number of correct choices per item. Scoring was difficult and students experienced doubt with the unequal number of choices per item. Third, students need some familiarity with the test format. One student was confused on the student-centered test and wrote sentences instead of definitions for part of it. Finally some control for test effect between the tests should be put in place. Items such as

*commute* and *make an effort* became progressively easier apparently because they were learned in the tests and subsequent interviews.

Given these limitations, the research questions are answered with caution. In response to RQ<sub>1</sub>, I have chosen to rank the sentence completion test as most practical in a class with close to 30 students or more, and the student-centered test as the most practical for smaller classes. The word association test is less practical than the other two.

Answering RQ<sub>2</sub> is more complex. The word association test measured meaning and use aspects equally even when these were not taught equally in the text. Thus, it did not meet the criteria for a fair assessment, and would be more suitable for a theoretical construct. The student-centered test was superior in the students' eyes and it gave them a chance to demonstrate their knowledge. And, in terms of aspects of word knowledge, it was easy to design the test to match the aspect I wanted. It was, however, less controllable for consistent grading and it did not measure production. The sentence completion test measured productive recall, but combined use and meaning in the prompts, making it difficult to isolate which aspect students were thinking about to arrive at the answer. If the teacher and students wanted to know which aspects of a word to focus on, this test may not be as informative.

In terms of validity for the purpose of informing me and my students of their current knowledge of the vocabulary items, the student-centered test appeared to be more valid than the sentence-completion test, which was in turn more valid than the word association test. If a teacher in a similar context were to choose a test based on practicality and validity, I would recommend the student-centered test especially in smaller classes. However, if the teacher specifically wanted to measure productive recall, or had a large class, I would recommend the sentence completion test. This and future studies investigating short-term achievement tests can help teachers decide which test method is most informative and practical for their course.

---

## References

- Harrison, G. (2007). *University issues*. Mishima, Japan: Unpublished textbook manuscript.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge UP.
- Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogards & B. Laufer (Eds.), *Vocabulary in a second language: Selection acquisition, and testing* (pp. 209-227). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- West, M. (1953). *A general service list of English words*. London: Longmans, Green and Company.
- Xue, G., & Nation, I. S. P. (1984). A university word list. *Language learning and communication*, 3(2), 215-229.

## Appendix

Table 5. Aspects of Word Knowledge (based on Nation, 2001) of Primary Focus in the Textbook (Excerpt)

Item	Form			Meaning		
	spoken	written	word parts	form & meaning	concept & referents	associations
complicated					R P (relationships)	
send the wrong message	R	R			R P (mis-understanding)	
relate to...	R	R		(LW)	R P	
not that great a...	R	R			R P	
make an effort						
gossiping				R P		
put sb down	R	R			R P	R P
not take long to...	R	R		R P	R P	
make fun of	R	R		R P	R P	R P
lead sb to	R	R			R P	
tease				R P		
get to sb				R P		
take on...	R	R			R P	

Item	Use		
	grammatical functions	collocations [ & MWU]	constraints on use
complicated	R P has got	R P things	
send the wrong message	R P sb...by doing st	R P	
relate to...	R P to st/sb		
not that great a...	R P a	R P	R P
make an effort		R P	
gossiping			
put sb down	R P put sb down		
not take long to...		R P start / realize	
make fun of		R P	
lead sb to	R P sb to do st		
tease			
get to sb		R P	
take on...		R P too much work	

R = receptive knowledge, P = productive knowledge, LW = loan word used in Japanese (not necessarily with same collocations)

Table 6. Test Results

Test	Student	1	2	3	4	5	6	7	8	9	10	11	12	total	percent
		complicated	wrong message	send the effort	make an gossip	not take long	tease	get to	downtime	stress	commute	I hear what you're saying	stuff		
Sentence Completion	Y	1	1	0	1	1	1	1	0	1	0.5	0	0	7.5	62.50
	H	1	1	0	0	0	1	0	0	1	0	0	0	4	33.33
Student-Centered	Y	1	1	1	1	1	1	1	1	0	1	1	1	11	91.67
	H	1	1	0.5	0.5	0.5	0.5	.5	0	0.5	0.5	0	0	5.5	45.83
Word Association	Y	0.67	1	1	1	0.67	1	0.5	1	1	1	0.75	0.67	10.26	85.50
	H	0.67	0.25	1	0.67	0.5	1	0	0.5	1	0.67	0.5	0.33	7.09	59.08

$M_{\text{sentence-completion}} = 47.92$ ;  $M_{\text{student-centered}} = 68.75$ ;  $M_{\text{word association}} = 72.29$

---

---

## Vocabulary Test (student-centered test I)

Follow the directions for each part. You will be asked to either write a sentence, a definition, or collocations.

I. For the following words or multi-words, write a sentence. Try to show that you understand its meaning and how the word is used.

1. complicated

---

2. make an effort

---

3. not take long to

---

4. downtime

---

5. stuff

---

II. For each of the following, write the meaning.

6. send (somebody) the wrong message

---

7. gossiping

---

8. tease

---

9. get to (somebody)

---

10. (a) commute

---

11. I hear you

---

III. For the following word, write at least one collocation.

12. stress

---

### Sentence Completion Test (excerpt)

- Relationships were simple at first, but lately things have gotten c (complicated).
- The American high school teacher never touches students because she doesn't want to send them the wrong m (message).
- I'm not working hard. I need to make more of an e (effort).
- We know it's not polite to talk about other people's weak points, but we still g (gossip).
- It didn't t (take) l (long) for people to notice I'm from another part of Japan.

### Word Associations Test (excerpt)

Directions

Below each vocabulary word are two boxes. The left box has at least one choice that relates to the meaning of the vocabulary word. The right box has at least one choice that matches the grammar or collocation of the vocabulary word. Each box has at least one correct answer. Circle all the correct associations with the vocabulary word..

Example: exhausted

smelly	very tired
used up	dirty

I'm...	...mood
That train ride was...	movie was...

1. complicated

confusing	surrounded
strange	curious

relationship	freely
has gotten	question

2. message

plan	misunderstanding
answer	opinion

talk...	forget to have a...
send the wrong...	teach a...

3. effort

attempt	product
deal	note

make an...	plan an...
try an...	take an...

4. gossip

share	laugh at
complain	talk about

lies	party
kindness	light

5. not take long

quickly happen	count up
suddenly do	connect to

...for them to see	...to realize
...to apologize	...to effort