

# Curriculum Design and Vocabulary Levels of Japanese Returnee Students: Equivalence of Vocabulary Test Forms

David Aline

## INTRODUCTION

The increase in Japanese families living abroad as a result of companies transferring employees overseas has brought a unique situation to the teaching of English in Japan. As families return to Japan, the children reenter the Japanese educational system. Often the children's English has improved to a great extent as a result of living in a country where English is the native language, or living in a foreign country but attending an international school where English is the language used for instruction. The term "returnee" has been applied to such students when they return to Japan.

As the number of returnee students has increased, many Japanese universities have modified their curriculum to include special returnee classes. Rather than requiring a returnee student to remain in an English class within their department with students of a lower proficiency level, the returnee classes enable many students at similar levels of proficiency to study together and, therefore, to improve their English in an academically challenging environment. However, little research has investigated the importance of an appropriate curriculum for returnee students.

## THE STUDY

Numerous areas of research could assist in the design of curriculum and choice of materials for returnee courses. This study

examines returnee students' vocabulary levels for the purpose of selecting appropriate materials for the greatest learning in return for the time invested in study. Specifically, the study tests returnees' vocabulary levels and explores the relationship of vocabulary levels to curriculum design of returnee courses.

### Literature Review

Research into the learning of vocabulary has demonstrated the importance of students' vocabulary levels in the selection of course materials. Most of this research has centered on levels needed to comprehend reading materials as opposed to spoken language (Hirsh & Nation, 1992; Laufer, 1987, 1989; Laufer & Sim, 1985). Laufer (1989) and Liu and Nation (1985) demonstrated the need for second language learners to understand at least 95% of the vocabulary in a text in order to achieve satisfactory comprehension and to be able to guess new words from context. If returnee students are to be able to guess words from context — one important learning strategy, especially for assisting learners gain autonomy — and comprehend the text — a prerequisite for language acquisition in psycholinguistic theories of language learning stressing the importance of meaning in language learning (Krashen, 1985) — then they will best be assisted in this process by materials that are within their vocabulary levels. If a returnee's vocabulary level covers 95% of a text, that learner will have a better chance of learning new vocabulary by guessing and acquiring new language through meaningful input.

### Statement of Purpose

The design of this study is twofold. First, the study looks at the vocabulary levels of returnee students studying in university. The discussion section relates those levels to methodology for teaching vocabulary in special classes for returnees. Therefore, the following research question was investigated:

1. What is the receptive vocabulary level of returnee students in a university course for returnees according to a

vocabulary recognition test?

Second, the study examined the equivalency of four forms of a vocabulary levels test. The research question investigated was:

2. Are the four forms of the vocabulary levels tests equivalent for the population of returnee students in a Japanese university?

## METHOD

The data collection of this study consisted of the administration of the four forms of the vocabulary levels test and collection of biographical data.

### Subjects

The subjects participating in the research were 112 students in two intact classes at a major Japanese university. For the purposes of this study, a returnee student was defined as any student accepted into the returnee English course. This includes students who have never lived overseas. The course is open to all returnee students and students who feel their English level is high enough to succeed in the course. Some students who are too advanced for their regular English classes are given letters of recommendation to the course from their regular English professors. All students must pass an entrance examination to be allowed to register for the returnees' English course. For the classes involved in this study, the entrance test involved watching a video of a news interview with a professor and summarizing the main points of the interview. The teacher then checked the essay for comprehension of the primary information in the video. The number of years that the student has lived overseas is also considered in borderline cases. The video is rather difficult; some students are not permitted to register for the class. Only 6 students of the 112 students in the returnee classes in this study had never lived overseas. The average number of years having lived overseas was 5.78.

There are two reasons for using in this study a definition of a

returnee as any student in the returnee courses. First, returnees have such a variety of experiences that they are difficult to define. Is one year overseas enough to be a returnee, or must one have lived at least three years overseas? Returnees have lived in a variety of countries, not just countries where English is the main language used. Returnees have attended a variety of types of schools, started living overseas at different times, and returned to Japan for short periods to again move overseas. Second, ecological validity is central to this study. This study seeks to evaluate the vocabulary of students in returnee classes so that a curriculum can be designed to meet their needs. Therefore, if all students in a returnees' class are not included in the data collection, the results will not reflect the needs of all the students in the class but only those within the range of a strict definition. This study is therefore focused on students in a returnees' course rather than only on returnees, however defined. The average years spent overseas of 5.78 should be enough to demonstrate the nature of the students as a whole as having lived overseas extensively.

The average age of the students was 19.8. They had lived overseas for an average of 5.78 years, had started to learn English at the average age of 9.84, and had been learning English for an average of 9.24 years. 86 of the subjects were women and 26 were men. Two subjects were lost due to mortality; they dropped the class. Their data was not included in the final analyses. A total of 112 subjects completed all four forms of the vocabulary levels test and the biographical information.

### Materials

A questionnaire was used to collect biographical data on the subjects in this study. Information included: age, years lived overseas, age of starting to learn English, total number of years learning English, and age at which one began to learn English.

The four forms of the vocabulary levels test were supplied by Paul Nation (See Nation (1983, 1990) for a sample of form A of the

vocabulary levels test and in-depth explanation of test construction). Each of the four forms (A,B,C,D) contains five sections: 2,000 word level, 3,000 word level, 5,000 word level, 10,000 word level, and University word list level. Each form has 90 items testing 180 words. Each section has 18 items testing 36 words. The test designer explains the rationale behind this: “Although only 18 words are matched at each level, in fact 36 words at that level are tested. This is because the distractors in the test are not meanings but words” (Nation, 1990, p. 261). However, only 18 items are scored per level. Each section has six groups with each group testing six words through three definitions. For example, the first group of the 2,000 level test of form A looks like this (Nation, 1990, p. 265):

1	original	
2	private	_____ first
3	royal	_____ not public
4	slow	_____ all added together
5	sorry	
6	total	

The students are to match the vocabulary items on the left with the definitions on the right by writing the correct number in the blank.

The level sections of the test are based on frequency counts of words in English. *A General Service List of English Words* (West, 1953), *The Teacher's Word Book of 30,000 Words* (Thorndike and Lorge, 1944), and *A Computational Analysis of Present-Day American English* (Kucera & Francis, 1967) were used as a basis for selecting the words to be tested in each level section of the test. For example, the 2,000 word level section has words chosen from the most frequently occurring 2,000 words in the English language. To learn these words provides learners with greater return for the time they invest in learning because the learners are more likely to encounter them in the language input they receive. Conversely, the more often a word

occurs, the likelier it is that the word will have been learned. The results of the test can then serve as a guide for teaching. In the words of the test designer, "The basic idea behind the vocabulary test is that the statistical distribution of vocabulary should guide the teaching and learning strategies" (Nation, 1990, p. 263).

The University word level section is unique in that it was constructed using the University Word List rather than a general frequency count of English. The University Word List (Xue & Nation, 1984) is based on a frequency count of words in university textbooks. This list spans the other word lists from about the 3,000 word level to the 10,000 word level. Nation (1990) suggests that students planning to attend institutions of higher education where English is the language of instruction would benefit most from a curriculum focused on these words.

#### Procedures

Biographical data were collected as part of the normal class procedure at the beginning of the term. Students worked in pairs, interviewing their partner and filling in the questionnaire with their partner's biographical data.

The four forms of the vocabulary levels test were administered over a four-week period during the normal class session. The subjects were randomized and the forms were counterbalanced. The time of the test administration within the class was also randomized and counterbalanced; the first and third administrations were conducted at the beginning of the 90 minute class sessions and the second and fourth at the end. The teacher gave instructions for the test; an example written on the blackboard was used to clarify procedure. Subjects were given as much time as they required to complete the test. Nation (1990, p. 261) suggests a maximum time of 50 minutes be allotted for sitting one form. All subjects completed the test well within 30 minutes. Upon completion of the test, students began work on their regular classroom assignments.

The tests were scored for the total scores for each of the four

forms and the total scores for each of the level sections within each form. The total scores within each section are more important for understanding a student's vocabulary level than the total form score because "If someone scores 12 or less out of 18 in a section of the test, then it is worth helping that learner study the vocabulary at that level" (Nation, 1994, p. 262). Since this test focuses on levels, the total score will not help the teacher to understand a particular student's level. Rather, the score within each level can aid the teacher in deciding the vocabulary level at which the student would most benefit from further study.

### Analyses

Statistical analyses of the data were conducted using SPSS software. The final scores of each section of each of the four forms along with the total scores for each form and the biographical data of gender, age, number of years lived overseas, number of years of English learning, and age of starting to learn English were correlated using Pearson product-moment correlation. The four assumptions underlying Pearson product-moment correlation were checked to see if this statistic could be used correctly for these analyses (Brown, 1988, pp. 136-139; Hatch & Lazaraton, 1991, pp. 436-437).

Since the tests were counterbalanced, and the total scores on test forms were not correlated with subtest scores, the tests met the first assumption of independence.

Analysis of the descriptive statistics shows the success of meeting the second assumption of normal distribution. Tables 1 through 4 display the descriptive statistics for the subtests of each of the four forms. Each subtest had a possible score of 18. As can be expected with the subtests of levels tests, the subtests below the vocabulary levels of the subjects were negatively skewed. Consequently, the negatively skewed subtests, 2,000 and 3,000-word levels, were removed from subsequent analyses because they did not meet the assumption of normal distribution.

Table 1  
Descriptive Statistics for Form A

	2,000	3,000	5,000	10,000	U.	Total
Mean	17.01	16.9	15	7.82	10.9	67.6
STD	1.07	1.48	2.1	3.39	3.3	8.71

Table 2  
Descriptive Statistics for Form B

	2,000	3,000	5,000	10,000	U.	Total
Mean	16.9	16	13.6	8.15	13.1	67.7
STD	1.07	1.79	2.98	3.3	3.1	9.21

Table 3  
Descriptive Statistics for Form C

	2,000	3,000	5,000	10,000	U.	Total
Mean	17.03	16.5	13.2	8.27	14.4	69.4
STD	1.33	1.51	2.63	2.87	2.67	8.2

Table 4  
Descriptive Statistics for Form D

	2,000	3,000	5,000	10,000	U.	Total
Mean	17.8	16	14.3	7.96	14.9	70.9
STD	0.49	2.09	2.52	3.14	2.5	7.75

The 10,000-word level subtests on all forms are normally distributed and can confidently be analyzed further. The 5,000-word level subtests and the University Word List level subtests are slightly negatively skewed; two standard deviations above the mean go just slightly beyond the maximum score of 18. These subtests will be included in further analysis. However, caution is suggested in drawing conclusions about equivalency on these subtests for this population.

The third assumption, the variables measured are on an interval scale, was met since the test scores and biographical data are all interval scales except for gender.



The fourth assumption is that the sets of scores are linear; analysis of scatterplots showed linearity.

For the analysis of equivalency of the four forms, the means were “compared to determine their equivalence,” and the correlations computed and then “interpreted as an indicator of the equivalence” (Bachman, 1990, p. 183).

## RESULTS

### Vocabulary Levels of Returnee Students

The results of the average scores on the four forms of the vocabulary levels subtests are presented in Tables 1 through 4.

Evaluation of the results will follow the test writer’s suggestions. Nation (1990) suggests that “If someone scores 12 or less out of 18 in a section of the test, then it is worth helping that learner study the vocabulary at that level”...“and learners’ scores on the test can be taken as a close approximation to the proportion of words in the test that they know. A score of 12 out of 18 indicates that approximately one-third of the words at that level are not known” (p. 262). As is clear from the mean scores in Tables 1 through 4, the returnee students scored fairly close to perfect on the 2,000 and 3,000-word level subtests and, therefore, can recognize the meanings of the first 3,000 most frequent words of English. On the 5,000-word level subtests, the returnees displayed knowledge of a good three-fourths of the words. Thus, they could still benefit from some focus on the words at this level. However, Nation (personal communication, July 26, 1995) expressed the opinion that the score on this level is impressive enough that the students would not need to study much at the 5,000 word level. How this focus might be accomplished will be taken up in the discussion section of this paper. At the 10,000-word level the subjects were on average unable to score higher than 8 out of 18; they are obviously not at the 10,000 word level yet. With an average score of 14 on the 5,000-word level subtests, and no other subtests for the levels between 5,000 and

10,000, it is safe to say that the returnee students, on average, are at or above the 5,000-word level. In addition, the University Word List level subtest showed the returnees scoring an average of 13 out of 18 on the four forms. As explained earlier, the University Word List level subtest tests words from a frequency count of university textbooks. These words range in overall English frequency from the 3,000-word level to the 10,000-word level. However, why the subjects scored so high on this level needs some further explanation. The students may be using some English textbooks or terms for other university classes beyond their English language classes. Or, since “many words in the University Word List are of Latin derivation” (Nation, 1990, p. 262), they might have had specific training on Latin prefixes and roots used in English while living overseas. Another possibility is that many of these students are currently studying Latin based languages as their major in university. A significant number of the returnees are French, Spanish or Portuguese majors: 33 out of 112, or 34%. In order to test this possibility, a multiple analysis of variance (MANOVA) was conducted on the University Word List level subtest and the dependent variable of majoring in a Latin based language or not. The significance of F was .053, which is not statistically significant at the  $p < .05$  level. Since it appeared to be close to significant, I decided to conduct separate T-tests on each of the four forms of the University Word List subtests. Only form C was significant with an F value of .002. Although the T-tests are not correct statistical procedure, and analysis should have stopped with the MANOVA, these results do raise some questions that should receive further research attention. Especially since only the students’ majors were considered, whereas many returnees reported French or Spanish as a third language.

#### Reliability of Vocabulary Levels Test Forms

The reliability coefficients are reported in Table 5 for each of the forms’ subtests and totals. The reliability coefficients for the 2,000-word level subtests are very low because the data was skewed.

Table 5  
Cronbach's Alpha Reliability for Vocabulary Levels Test Forms

	2,000	3,000	5,000	10,000	U.	Total
Form A	.3082	.6017	.6040	.7010	.7012	.8617
Form B	.2428	.6024	.7061	.7092	.7299	.8705
Form C	.5384	.5332	.6324	.5865	.6709	.8481
Form D	.1145	.6600	.6614	.6632	.6938	.8417

The total reliability coefficients for each form are at a good level. The other reliability coefficients are not too low but not very high. With a test of only 18 items, this is to be expected. If higher reliability is desired, the forms can be combined. However, this increases the time needed for administration. When used for research purposes, the researcher will want to increase reliability by combining forms. On the other hand, the teacher in the classroom will wish to conserve time by using single forms and devote the time saved to the study of vocabulary.

#### Biographical Data Correlations

The biographical data of the returnee students were correlated together with the total scores on all subtests of the four forms. Since the biographical data is not central to this study as it is used mostly to define the English learning experiences of returnee students, all correlations will not be presented in table form; rather, only the significant correlations will be briefly discussed. Gender correlated only with the 10,000-word level subtest of Form A and Form B at -.3389 and -.2628, respectively. Age did not correlate significantly with any other variables. Reported number of years of studying English correlated with age at which English learning began and number of years spent living overseas, -.8263 and .2368 respectively. The number of years of English study or learning also correlated positively with all four forms of the 10,000-word level subtest, the 5,000-word level subtests of Forms C and D, and the 3,000-word level subtest of Form A. Clearly, the greater number of years of

study will produce higher vocabulary levels. Age at the start of learning English correlated negatively with the number of years lived overseas,  $-.3429$ . For the levels test, age at start correlated positively with the 10,000-word level of all four forms, but not with any other level. Again, perhaps this is a result of more years studied producing higher vocabulary levels. Years lived overseas correlated only with Form C's 10,000-word level subtest. Since these variables were not the main focus of this study, and since correlations do not show a cause and effect relationship, further research will need to be conducted with the interesting variables in these results as a basis of the research questions.

#### Equivalency of Vocabulary Levels Test Forms

The equivalency of the four forms of the Vocabulary Levels Test was checked through a comparison of means and Pearson product-moment correlation. The 2,000-word level and 3,000-word level subtests were included in the correlation but will not be discussed for equivalency because of their skewed distributions. In addition, all variables were correlated, however, because of the large size of the correlation matrix, and for clarity, the results will be reported in individual tables. Significance level for all correlations was set at  $p < .01$ .

The means and standard deviations of the 5,000-word level subtests were compared and found to be somewhat similar (See Tables 1 through 4). The correlations, from the overall correlation matrix, of the 5,000-word level subtests from the four forms are reported in Table 6.

Table 6

#### Correlations of the 5,000-word Level Subtests

	Form A	Form B	Form C
Form B	.5503*		
Form C	.5557*	.5433*	
Form D	.4914*	.6208*	.5613*

\* $p < .01$

The correlation coefficients for the 5,000-word level subtests show that all the forms correlate significantly. Therefore, it can be stated that these forms are not different. However, the correlations only explain a minimum of 24% and a maximum of 39% of the variance between the test forms. Combining Forms A and B, and correlating them with the combined scores of Forms C and D, gives a significant correlation of .7137 which explains 51% of the variance. Hatch and Lazaraton (1991) suggest a higher correlation is needed, "This is a fairly strong correlation; the overlap is 50.4%. However, if you hoped to show that the two tests measured basically the same thing, the correlation isn't very strong. You would want a  $r$  in the high .80s or .90s. You would want a very high overlap of the two measures" (p. 441). Still, depending on how the test is to be used, it appears that these forms share a "fairly strong correlation."

Table 7 presents the correlations for the University Word List level subtests. The results here are similar to the 5,000-word level subtests. All forms correlate significantly with each other, but not to an extremely great degree. It can still be said that these forms are not very different from one another.

Table 7

Correlations of the University Word List Level Subtests

	Form A	Form B	Form C
Form B	.5067*		
Form C	.5686*	.5545*	
Form D	.5097*	.4181*	.5024*

\* $p$ , .01

For the 10,000-word level subtests, the subtests with the most normal distribution, the correlations are again all positively significant (Table 8). The forms are not very different.

Table 8  
Correlations of the 10,000-word Level Subtests

	Form A	Form B	Form C
Form B	.6782*		
Form C	.5875*	.5415*	
Form D	.4712*	.4466*	.4748*

\* $p < .01$

Furthermore, the correlations are also not as high as might be hoped. However, Forms A and B do have a higher shared variance of 46%, while Form D has the lowest. Unfortunately, the same pattern does not hold for the other levels. Therefore, it is difficult to say that any one form needs revision. These correlations also refer only to the population of Japanese returnee students in this study. Further equivalence of forms awaits future research with other populations.

From the statistics analyzed above, it appears that the forms are not different and could be used in research as equivalent forms, or used by teachers to assess long term vocabulary gains. The forms can certainly be utilized reliably in assessing students' current vocabulary levels in a counterbalanced testing administration to ensure against students looking at other students' test papers.

## DISCUSSION

The returnee students in this study were able to recognize 75% of the 5,000 most frequent words in English and 75% of the University Word List. They would clearly benefit most from study of vocabulary at these levels since these words occur with a greater frequency than lower frequency words and would, therefore, be more likely to be met in future reading. However, studying words from the University Word List will depend on the students future educational plans and current textbook reading requirements in their current university courses. Since some of the students in this study

have already intimated their interest in applying for graduate programs overseas, and some are actually in the process of applying, studying words from the University Word List will improve their comprehension of university textbooks. Also, students who are currently studying in courses that use English language university texts would benefit from focusing on this list as it would also increase their comprehension for their course work. They would also benefit if they use English texts as resources for term papers or theses.

Nation (1990, pp. 262-263) suggests some ways to assist students to learn vocabulary at each level. At the 5,000-word level, he suggests the following:

1. Training in guessing words in context
2. Wide general reading — novels, newspapers, university texts, etc.
3. Intensive reading of a variety of texts

Training the students to guess words in context will be beneficial if the students do not already use this strategy. Otherwise, it could be a waste of time. The teacher might want to make sure that the time spent in class will be teaching a new strategy, or that the review will be helpful. The current curriculum of the returnees' course utilizes a wide variety of general readings, including: magazine and newspaper articles on a variety of topics, short stories, and some excerpts from university texts. Hwang and Nation (1989) found that reading newspaper articles on the same topic helped to increase the students' vocabulary because the vocabulary related to that topic was repeated across articles. Therefore, the students were exposed to the same vocabulary a number of times and in a variety of grammatical contexts. The intensive reading of a variety of topics from magazine articles in the returnees' course helps by exposing the students to a frequent repetition of words used in this genre. In order to facilitate the students' acquisition of vocabulary at the 5,000-word level through the use of magazine articles, the article can be scanned into a computer and analyzed with word frequency count software. The

list of words from the article could then be compared with the 5,000-word list, and a list of words from the article falling within the 5,000 most frequent words of English could be drawn up. These words could then be pretaught to the students so as to increase their comprehension of the article during initial reading. The preteaching might also have the effect of focusing the students on the collocation of the words. Additionally, the collocations might be used as a preteaching technique. The collocations could also be found with computer software designed for this purpose.

For the University Word List level, Nation offers these suggestions:

1. Learning words in the University Word List
2. Intensive reading of university texts
3. Learning prefixes and roots

The returnee students who need to study in this area could be shown the University Word List (Nation, 1990, pp. 235-239) and assisted with learning strategies for vocabulary development. The words in the list are numbered for frequency and range. Thus, students could concentrate on the most frequent words for the most benefit, or focus on the frequency range that suits their particular level. Many of my returnee students do not seem to use the strategy of guessing a word's meaning from its prefixes and roots. This is a common technique for first language learners in English. Further research is needed so as to understand the extent of the use or knowledge of this technique among Japanese students, possibly using the think-aloud research technique (Nunan, 1992, p. 117), and to further knowledge into techniques for teaching prefixes and roots and the learning outcome of different techniques.

On the 10,000-word level Nation recommends "Activities similar to the 5,000-word level, combined with learning prefixes and roots" (p. 263). However, since the average score on this level of the test was around 8 out of a possible 18 for these subjects, the returnees might benefit more from concentrating on achieving a higher score



at the 5,000-word level. On the other hand, those students who scored higher on the 5,000-word level test will profit equally from intensive reading, wide general reading, and guessing strategies.

### CONCLUSION

This study has looked at the vocabulary level of students in a returnees' course. The results have shown that the returnees have a recognition knowledge at or above the 5,000-word level. Suggestions for teaching have been considered. Additionally, the four forms of the Vocabulary Levels Test have been found to be statistically not different for the 5,000-word level, 10,000-word level, and University Word List level subtests.

### References

- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Brown, J. D. (1988). *Understanding research in second language learning*. Cambridge: Cambridge University Press.
- Hatch, E. & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. New York: Newbury House.
- Hirsh, D., & Nation, I. S. P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8(2), 689-696.
- Hwang, K. & Nation, I. S. P. (1989). Reducing the vocabulary load and encouraging vocabulary learning through reading newspapers. Unpublished paper.
- Krashen, S. (1985). *The input hypothesis: Issues and implications*. London: Longman.
- Kucera, H., & Francis, W. (1967). *A computational analysis of present-day American English*. Providence, RI: Brown University Press
- Laufer, B. (1987). The lexical perspective of reading comprehension. *English Teachers Journal* (Israel), 35, 58-67.
- Laufer, B. (1989). Why are some words more difficult than others? — Some intralexical factors that affect the learning of words. *IRAL*, 27,3.
- Laufer, B., & Sim, D. (1985). Measuring and explaining the reading

- threshold needed for English for academic purposes texts. *Foreign Language Annals*, 18(5), 405-411.
- Liu, N., & Nation, I. S. P. (1985). Factors affecting guessing vocabulary words in context. *RELC Journal*, 16(1), 33-42.
- Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines*, 5(1), 12-25.
- Nation, P. (1990). *Teaching and learning vocabulary*. Boston: Heinle & Heinle Publishers.
- Nunan, D. (1992). *Research methods in language learning*. Cambridge: Cambridge University Press.
- Thorndike, E., & Lorge, I. (1944). *The teacher's word book of 30,000 word*. New York: Teachers College, Columbia University.
- West, M. (1953). *A general service list of English words*. London: Longman, Green and Co.
- Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication*, 3(2), 215-229.