

原 著

Designing and Testing a User-Friendly Vocabulary Knowledge Scale for Teaching Essential English Verbs and Adjectives to Asian Students

John Paul Loucky

<Abstract>

This study aimed to examine some of the most important factors involved in learning new words by designing a user-friendly Vocabulary Knowledge Scale (VKS). This Scale helped assess and monitor change in vocabulary knowledge among Asian college students when it was used as an instructional aid in learning important adjectives and essential verbs. The instrument proved to be very effective and reliable in helping various Chinese, Japanese and Korean college students learn how to better attend to, assess, activate and acquire important English adjectives and verbs.

Key words: Asian students' English vocabulary levels, Vocabulary Knowledge Scales, lexical competence, second language vocabulary acquisition (SLVA); measuring aspects of vocabulary knowledge

1. Introduction and Statement of the Problem

Increasingly teachers and researchers are recognizing what most ESL/EFL learners have sensed personally all along, namely that "lexical competence is at the heart of communicative competence," as Meara stated (1995, p.35). While recognizing this fact, how to best measure and evaluate lexical acquisition has nevertheless been a thorny issue. This is because, as Lessard-Clouston (2000) noted, "Vocabulary knowledge is a complex construct, and assessing vocabulary knowledge is thus a complicated task (Read, 2000)."

Chinese, Japanese, and Korean students may be able to produce L1 translations, but frequently can give few L2 synonyms and generate even fewer sentences using target terms with appropriate grammar and collocations, most probably due to being trained by the traditional

direct translation (*yaku-doku* in Japanese) method. Such students have usually only learned translation, but possess knowledge of very few actual target language synonyms or vocabulary words that they can use productively in their own oral or written expression. Thus they are trained to be receptive rather than productive language learners, accustomed to translating everything into their native tongue rather than learning to think in the target language. Likewise such students are rarely asked to move much receptive vocabulary knowledge to a potential or real active production level.

In many kinds of VKSs, when a student says: "I think it means _____," or "It means _____" they are at least USING the target language word in a definition sentence correctly, although not generatively in an original sentence of their own. Waring's point (2000, p. 12) is that such a scale may give "insufficient evidence of the depth of

Associate Professor, Early Childhood Education and Care Department

knowledge of that ability” or word. Read (1997) also concluded that a rating scale would require much refining to improve its validity.

In response to such criticism, this new form of VKS gives much information that is useful to teachers for planning instruction, as well as to learners for their study and activation of new vocabulary. Whereas Paribakht and Wesche’s original VKS has a 5-point scale that combines “self-report and performance items to elicit self-perceived and demonstrated knowledge of specific words in written form” (1996), L1 and L2 are not distinct, nor are perception versus production. By contrast, this writer’s Evaluator provides researchers with a sharper instrument that can quickly elicit clearer data on more distinct, detailed aspects of each learner’s vocabulary knowledge. This assessment tool practically helped Chinese, Japanese and Korean college students alike learn how to better attend to, assess, track and acquire important English adjectives and verbs.

2. Literature Background

2.1 Testing Aspects of Vocabulary Knowledge

One must be clear on what aspects of vocabulary knowledge one wishes to measure and teach. Both Lessard-Clouston (2000) and Nation (1990; 1994; 2001) provide excellent surveys of this field. As Lessard-Clouston (2000) noted,

Much of the recent work on L2 lexis focuses on vocabulary size, growth, and use, and in a framework relating these different aspects of vocabulary Nation (1993) makes a number of helpful points. First, one’s ‘skill in language use depends on vocabulary size,’ and one should thus be familiar with high frequency words and ‘the general academic vocabulary that is common in many academic disciplines (Nation, 1993, pp. 118 & 120)... This framework by Nation (1993) makes clear that there are various types of words and also suggests that there are diverse types

of vocabulary knowledge... two particular aspects of vocabulary knowledge discussed in the research literature are especially relevant to the present study: vocabulary breadth and depth. Breadth vocabulary knowledge indicates something of a person’s vocabulary size, or approximately how many words one knows. In contrast, depth vocabulary knowledge concerns the quality of a person’s knowledge of a word—how well someone knows a specific word or set of words. Wesche and Paribakht (1996) have provided an overview of a range of test options for assessing breadth and depth vocabulary knowledge... (Lessard-Clouston, 2000, pp. 1 & 2)

More instructors are recognizing the clear connection between learners’ strategy use-awareness and their language proficiency levels (Tsudajuku, 1992; Yamato, 1997; and Kinoshita, 2003). Similarly, the important connection between a language learner’s vocabulary level in particular, and that individual’s subsequent reading and listening comprehension ability in the target language (TL) has also long been known. A further obvious reason to investigate students’ vocabulary levels is that they are most closely connected with learners’ comprehension ability. Alderson (2000, p. 35) states,

Measures of readers’ vocabulary knowledge routinely correlate highly with measures of reading comprehension, and are often, indeed, the single best predictor of text comprehension. Having to struggle with reading because of unknown words will obviously affect comprehension and take the pleasure out of reading. Research by Laufer (1989) and Liu & Nation (1985) shows that readers need to know 95% of the words in text to gain adequate comprehension and to be able to guess unknown words from context. Hirsh and Johnson estimate that in order to be familiar with 97% of the words in

text, a reader needs a vocabulary of roughly 5,000 words.

Recognizing these strong relationships makes such a study and comparison of vocabulary and language proficiency levels of prime importance, especially in Japan where few such studies can be found (besides those of Author, 1996; 1997a; 2002d, 2003a).

Finally, because of the central importance of developing learners' TL vocabulary, the need to better assess and focus on specific essential core lexical items and phrases has been clarified by Guest (2000). His results showed that to better develop learners' lexical skills, "the choice of lexical items for analysis or study should not be left up to the individual learner, but rather deliberately and explicitly guided and monitored by teachers (p. 180)." While one would be wise to allow for at least some learner autonomy, teachers should try to help students to practice vocabulary learning strategies and comprehension skills essential for more systematic lexical processing. These can in turn become a strong foundation for most other dimensions of second language acquisition. CALL multimedia can now help to better facilitate language development using learners' individual preferred learning styles and strategies (as shown at this writer's <www.CALL4ALL.us> website World CALL Directory).

For more than one decade the author has annually examined vocabulary levels and tests given in Japan (See Author, 1994-2005 of foreign and bilingual students as well as Japanese, especially to help determine college students' general base-line levels upon entrance to college, and their improvement rates after 1-2 years of college study. He has also contrasted these with the reading levels found in texts used on many college entrance tests in Japan (Brown and Yamashita, 1995; Author, 1996). Three types of studies were done to check on the usefulness of this new type of VKS called a Dual Assessment Vocabulary Instructor-Evaluator (or DAVIE, as shown in Appendix A) from three perspectives to

better confirm its practicality by means of a sort of triangulation. These three types of Evaluator checks all compared student "Receptive Input" self-reports with actual teacher-assessed "Productive Output" assessments of its five categories. They included 1) the first 100 EAP word families from International Christian University's computer-generated "Recommended EAP List" (Mizoguchi, et al., 1992; found in Author, 1996, pp.322-89); 2) 211 words from JACET's English Usage Special Interest Group, administered using lists of important verbs and adjectives; and 3) the DAVIE was used to help chart movement between its five categories of word knowledge for twenty essential verbs and twenty adjectives from within these larger lists. To help our foreign language learners acquire this most essential common core general and academic vocabulary, we must first assess what kinds of vocabulary knowledge our students have versus words they still most need to learn. In order to do so, several Vocabulary Knowledge Scales have been designed, such as those by Paribakht and Wesche (1993; 1996), and Zimmerman (1997). None of these had become very widely used as yet in Japan, therefore the author designed, tested and modified a VKS called a "Dual Assessment Vocabulary Instructor-Evaluator" for specific use with Asian students (see Author, 2002a, 2002d, 2005b). It is particularly useful in assessing learners' knowledge about particular domain areas or groups of words, both before and after instruction, as it clearly distinguishes between both L1 vs. L2 and receptive vs. productive aspects of vocabulary knowledge. While Zimmerman, Wesche and Paribakht have certainly contributed much to the field of testing vocabulary, more precise instruments have been needed that clearly separate various aspects of L1 and L2 vocabulary knowledge so as to gain a better picture of learners' respective developing mental lexicons, both L1 and L2. A simpler tool for in-class students' use which they can help to score and evaluate has long been needed, since few teachers have the time or training to analyze

longer, weighted VKS formats.

2.2 Improving Vocabulary Assessment by Developing a Student-Friendly Vocabulary Knowledge Scale

Rosszell (2007) examined “Extensive Reading and Intensive Vocabulary Study” in his recent dissertation. Among his recommendations for improving vocabulary teaching and learning was support for incorporating a more user-friendly Vocabulary Knowledge Scale, such as that of this writer. Although results of his own study clearly demonstrated how useful VKS can be as a research tool to monitor development of learners’ depth and breadth of lexical knowledge, he agrees with the author that rating of most traditional VKS is very time consuming. Recognizing this limitation of most Vocabulary Knowledge Scales, he states that

Loucky (2005) developed an adapted version [of a Vocabulary Knowledge Scale, called the DAVIE VKS] that his students were able to use on a regular basis to gauge their own lexical development. Using his system, learners initially assess their knowledge of words listed in a checklist, and the following week they demonstrate their knowledge of the words that they claimed they knew on the checklist. Author suggested a number of ways in which progress can be quickly measured. Although more research is necessary to verify the reliability and effectiveness of his approach, the development of a student-friendly version of the VKS would no doubt be useful not only in raising learners’ awareness of the notion of depth of lexical knowledge and in possibly having a positive washback effect, but in providing them with regular feedback on their progress and thereby motivating them to further develop their lexical knowledge. (p. 176)

A simpler word-knowledge checklist has also been tried (e.g., Horst, 2005; Horst & Meara,

1999; Zimmerman, 1997). A drawback of this type of test is that learners may over- or under-state their actual word knowledge, depending what they think teachers want to hear. So what is being tested is perceived rather than actual knowledge. Rosszell notes some of these common problems in self-report checklists of word knowledge, comparing the results of using several kinds of VKS. He found that

While Horst (2005) cites findings from Horst (2000) in which her learners were “able to provide accurate translation equivalents of about 80% of the words they rated ‘known’ (p. 365), that means that for the other 20% of the items, they had overestimated their knowledge on the checklist. As Read (2000) pointed out, ‘all that we can confidently say about a “yes” response to a word [on a checklist]...is that the learner is familiar with the word form and can identify it as a real English word’ (p. 148). (Rosszell, p. 95)

Using our own DAVIE VKS, and then comparing students’ self-perceptions or reported word knowledge with that which they can produce obtained similar results of 70-90% accuracy, varying mainly with their vocabulary proficiency, and being more accurate as their levels increase.

Zimmerman (1997) used a scale very similar to the VKS used in Rosszell’s (2007) study in her checklist (i.e., not requiring any knowledge to be demonstrated), nor providing any data on actual production of any of their claimed knowledge. In contrast to this, Rosszell reports that “Horst and Meara (1999) used a VKS-like scale on their checklist and found a 90% success rate in their single subject’s ability to provide ‘a correct or near correct translation equivalent’ (p. 315) for items that had been rated ‘definitely known’ on the checklist” (p. 95).

2.3 Learning How Lexical Knowledge Develops

There are various theoretical models of the complex process of second language vocabulary

acquisition (SLVA), but the aim of this book is to clearly describe a systematic practical approach to teaching and learning vocabulary and reading skills and strategies online, then illustrating it using some sample programs and websites.

In the process of reading or learning new words, guessing as to their meaning is simply too unreliable, and often leads to miscues/mis-readings and misunderstandings. Clearly, when our goal is to monitor growth in knowledge of individual words, checklists are simply too unreliable. Noting this, to increase the reliability of Zimmerman's results,

she eliminated from her study those learners who claimed to know two or more of the non-words in the checklist... Although instruments that require learners to demonstrate their knowledge take more time to complete and inevitably reduce the number of words that can be tested, by requiring learners to actually demonstrate their knowledge, many reliability and validity problems can be overcome... Despite Dolch's (1932) argument that testing depth of knowledge is impractical, given the incremental nature of vocabulary growth, such testing is essential ... if we hope to gain insight into how lexical knowledge develops. (Rosszell, p. 95-96).

The bottom line in determining the effectiveness of any particular approach to vocabulary development is clearly stated by Read (2004): "Ultimately the question is not about what learners know *about* a word but *what they can do with it: being able to pronounce it, recognize it in connected speech and writing, and use it fluently in their own production*" (p. 224, our emphasis).

2.4. Rationale for Testing Vocabulary

Why test language learners' vocabulary? Language teachers need to know what lexis their students already have acquired, and what they yet need to learn. Before testing, however, teachers and researchers need to clearly understand

why they are assessing their students and how the results will be used. Care must always be exercised when testing since it usually has a dual effect of not only producing information, but also influencing attitudes of both teachers and learners. How test results are used, therefore, could be either encouraging or discouraging, so we must try to be wise and careful in their use.

One may test each individual learner by means of several kinds of vocabulary tests including 1) Standardized reading tests, used to ascertain their approximate vocabulary and reading comprehension grade level relative to native reader norms; 2) Vocabulary Knowledge Scales (VKS), such as those developed by Wesche & Paribakht, (1996) and Zimmerman (1997), or using this Dual Assessment Vocabulary Instructor-Evaluator, designed specifically for use with Japanese students (Author, 2005b). 3) Finally Headword Level Tests (e.g., Nation, 1990; Laufer & Nation, 1995; Laufer, 1997) can be used to determine the approximate number of words known in the target language, either receptively, or productively.

Teachers need to ask several guiding questions while helping learners to meet new words, establish previously met vocabulary, enrich previously met vocabulary (through planned reencounters), develop vocabulary learning strategies, and develop fluency with known vocabulary. These helpful guiding questions are: a) What words do they need, b) How can they meet these new words, and c) How can we best test what they now know versus what they yet need to learn? (Nation, 1990; 1994).

The reason it is so important to help learners acquire the most frequent word families in English, according to Nation and Newton's calculations of text coverage (1997) is because knowing these will enable them to deal with 80% of the words they will usually meet in general written and aural texts. Learning another 570 of the most common core academic vocabulary will raise the potential textual coverage to about 90%. Thus as learners acquire more headwords and their reading grade

level rises, so will their level of comprehension. As Laufer (1997, p. 167) writes,

...the turning point of vocabulary size for reading comprehension is about 3,000 word families... The level at which good L1 readers can be expected to transfer their L1 reading strategies to L2 is 3,000 word families, or about 5,000 lexical items. Until they have reached this level, such transfer will be hampered by an insufficient knowledge of vocabulary.

She then goes on to explain how after that threshold level is reached, a student's degree of comprehension can be expected to increase by about 7% for each additional 1,000 words they learn. Indeed, by using such a conversion formula, Laufer (1997), Nation (1983), and Nation and Newton (1997) have shown us convincingly how indispensable having a good vocabulary is to good reading on repeated occasions, and that "Reading may be a psycholinguistic guessing game [in the words of Goodman, 1967], but words are the tools you need to play it right" (Laufer, 1997, p. 32).

A clear distinction between *receptive recognition* of vocabulary input on self-reports versus *productive recall* on output assessments may also be seen in this Dual Assessment Vocabulary Instructor-Evaluator's two formats. As Chappelle (1994, p. 163) has defined it, vocabulary ability includes both "knowledge of language and the ability to put language to use in context," or in Waring's (2000) terms, there are both *receptive understanding* as well as *productive use* aspects involved in "knowing" a word. This Evaluator can help us to better clarify them quickly for each student.

3. Materials Used

3.1 EAP and JACET Word Lists

First a DAVIE VKS was given to assess engineering students' knowledge of the first 100

word families from Mizuguchi et al.'s (1992) Recommended EAP List words. Then JACET (1995, 2000) English Usage SIG-selected adjectives and verbs were chosen to compare use of this new DAVIE VKS among Asian students. The EAP study by Mizoguchi et al. (1992) is probably the largest study of its kind done in Japan. The other two JACET studies gave special reference to corresponding synonymous verbs and adjectives in Japanese. The JACET English Usage SIG held a special symposium to overview and compare these synonym studies. Their research focused on 23 sets of synonymous verbs and 21 sets of synonymous adjectives in English. They set about to analyze in detail distinctive features of these words, according to semantics, collocations, and syntactic structures. Special attention was paid to

the differences between English verbs and their Japanese counterparts. The data are [sic] based on dictionaries, usage books, the COBUILD Corpus, and [was] supplemented by a special usage survey administered to ten informants of four English-speaking countries. The results of this research may be useful in the teaching of English as a second language or a foreign language. (JACET English Usage SIG, 2000, p. iv)

3.2. Designing a User-Friendly Dual Assessment Vocabulary Evaluator

Since effective language learning grows out of active, communicative use it is important for teachers to compare and assess both language learners' *receptive recognition or understanding vocabulary*, as well as their *active production or use vocabulary*. Thus the author designed this Dual Assessment Vocabulary Instructor-Evaluator to meet the real needs of Asian students and teachers. It has five aspects or states of vocabulary knowledge, ranging on this scale from categories A-E for rapid assessment, as well as having an optional 10-point productive rating scale. Word knowledge can move between these

states or categories, at times becoming dormant from disuse, as writing ability in Kanji characters often becomes.

Given first receptively, all DAVIE VKS items are equal in weight, so that only percentages for each column of a distinct type of vocabulary knowledge need to be calculated (e.g., For 10 words where all were known in Japanese, Column A would be 100%; if a learner can only give five L2 definitions that would be 50% under Column B; 3 words used correctly in sentences would give a Column C score of 30%). Then individual scores are recorded, and class averages figured. Next these are compared with productive scores, which later may be used again to measure each learner's degree of vocabulary growth, gain or accretion after instruction, even using the same Evaluator as a posttest instrument. These are its categories: *A. I know the Japanese meaning of this word; B. I know the English meaning; C. I can use this word in a sentence; D. I am not sure of its meaning, but think I've heard or read it in a phrase like _____; and E. I have no idea what this word means, it's completely unknown.*

When the test is given *objectively*, learners must give written (or oral, if given orally) *evidence of actual word knowledge* by giving L1 or L2 definitions for A & B, or by giving an entire sentence illustrating correct use of the word semantically. Although only item H gives a fully generative productive response, much can be learned by assessing a learner's verbal attempts to translate or define target language vocabulary in L1 and L2 as well. This scale also clearly distinguishes between ability to give L1 translations and L2 synonyms, which Paribakht and Wesche's (1993) VKS combines under one category (3), thereby confounding two linguistically distinct factors. One need not be able to use a word with syntactic precision (grammatical accuracy) to understand the word's basic meaning, although such ability does earn an extra point in overall scoring, both on this Dual Assessment Vocabulary Instructor-Evaluator's longer assessment form and also on Paribakht and

Wesche's VKS. While some may prefer to think of this more recent DAVIE as a VKS variation developed for the Japanese or Asian context, in many ways its design and use are distinct, as this study shows. When used without any weightings but just averaging each word knowledge category (A-E) using the Subjective Self-Report Form, students themselves can get a quick idea of their own vocabulary strength and become more actively involved in the learning process. The Objective Assessment Form must be evaluated by teachers, and Category A requires one that is bilingual, but advanced students may also be enlisted to help check L1 translations using dictionaries to check if needed. When this form is used without taking time for calculating any weighted scores for each category, the DAVIE becomes much more stream-lined, lighter version VKS, which is thus much easier to assess. Users themselves can also understand such a stripped down VKS much better being non-technical.

As these Asian students are taught to translate most of their language learning (learning by "direct translation" or *yaku doku* method in Japanese), it is naturally easiest for them to first be asked for a mother tongue translation of target terms, moving from simply checking off items on a self-reporting survey, then at a later date filling in a so-called "Objective Assessment" of familiar L1 translations (Category A), L2 (Category B) definitions, and familiar L2 collocations (Category D), and lastly moving to the more challenging whole sentence generative level (Category C). A test of the same target words assessing contextualized receptive knowledge can be made for any set of words, as shown below. This should be given only as a third stage of assessment, so additional knowledge gained from context clues and guessing based upon them does not interfere with testing of actual base-line vocabulary knowledge. This DAVIE VKS thus clearly separates its diagnosis of *contextualized receptive recognition* from its *decontextualized productive recall assessment* formats and times.

Testing *contextualized receptive understanding*

of target terms may be done using these words in single sentences or in a passage. Factors to consider are the learner's average reading level, the lexical density and transparency of a reading passage being tested, and what prior background knowledge is required to understand it. Since these factors can complicate testing, single sentences illustrating most common usage of target terms are preferred, as these would also be most frequently encountered in dictionaries. When using this DAVIE VKS, the most efficient way for teaching unknown terms is for teachers to provide illustrative sentences to students after assessing their responses to its objective form, stressing that they should focus on trying to guess the meanings of only those words which they missed to save time. As an example, an engineering class used the DAVIE to assess their knowledge of these ten words: *abandon, abbreviate, abide, ability, abnormal, abolish, abroad, absence, absolute, and absorb*. After averaging each student's responses and percentages of vocabulary knowledge per category, a student who missed three terms would be told to focus on sentences illustrating meanings of these missed words.

While similar in some ways to other Vocabulary Knowledge Scales (VKS) with five-item scales (See Read, 2000, pp. 132-138 for more on VKS), this Dual Assessment Vocabulary Instructor-Evaluator (also called DAVIE, or the Evaluator) has a simpler order, style and means of assessment that clearly distinguishes between L1 and L2 knowledge, and helps to trigger collocational knowledge for category D. Rather than simply spending 40% of test responses generating no measure of how unsure or unclear an item is, this type of question primes learners to guess a target term's meaning or collocations. This adds more knowledge to the scale, also training learners in trying to guess and remember new word meanings and common collocations, which are most essential language learning skills to develop.

Students are first asked to assess and calculate their own categories of vocabulary knowledge

for particular groups of words by totaling percentages for categories A-E. Later they show active or productive knowledge of words they checked off as known, including knowledge of A. L1 Translation equivalents, B. L2 Definitions, C. Common collocation phrase or whole sentence use. It is much less threatening for most Japanese students to begin by checking if they think they know a more familiar L1 Japanese translation, starting these self-reports based on their receptive knowledge or passive recognition first.

Each word may be assigned a value from 1-10 in order to calculate a weighted score showing one's depth of lexical knowledge about a target term or set of terms (as shown in Table 2, with 10 words times 10 points for relatively complete knowledge scores equaling 100%). This type of calculation is useful when one wishes to compare learners' L1 and L2 mental lexicons, as well as their receptive versus productive abilities, as a weighted score can help to better accentuate these different types of lexical knowledge and which aspects teachers and learners most need to work on. This study for example shows us that most Asian learners tested had little ability to define target terms, give common collocations or produce example sentences using them. Comparison of these distinct types of lexical knowledge may also help us to discover relationships between more passive versus active types of instruction and resulting weaknesses or strengths in language learners' development of their TL lexical processing abilities. In fact, there seems to be a lack of sufficient knowledge of the wide variety of vocabulary learning strategies available for use on the part of most Japanese students (Schmitt, 1997), who are generally trained to only translate, or rather to passively listen as teachers translate, having little or no chance to develop much practical, productive vocabulary through expressive use on their own. Active strategy training is needed.

An easier calculation of scores for this Evaluator is simply to check and compare productive use percentages with receptive self-

assessed claims for any gaps in vocabulary knowledge, while also arriving at an accuracy percentage. As Columns F-H in Table 4 clearly show, this DAVIE VKS instrument has had very acceptable levels of accuracy (from 79 to 92%) for four classes at different levels of English proficiency. This score may also be used as a *posttest evaluation*, which compares initial receptive self-reported vocabulary knowledge with actual productive knowledge after instruction to assess amount of vocabulary growth, either in terms of short- or long-term retention. On the initial test of this instrument with four classes at different levels, students' self reports had high levels of accuracy: 1) Japanese translations were 89.97% accurate, 2) English definitions were 86.75% accurate, and 3) English sentences were 89.5% accurate on average for these four classes, having 54 students whose vocabulary level averaged grade 3.6 relative to native U.S. norms (assessed by Riverside's Gates McGinite reading tests).

When this test is first given *receptively* it is just a self-reporting checklist to show levels of word knowledge. As such it was first used to assess students' knowledge of the first 100 word families on the *Recommended EAP Vocabulary List* (Mizoguchi, et al., 1992), which was a computer-generated list based on essential common core vocabulary found in college texts across ten academic disciplines, minus required junior high words (only 507 words in Japan at that time, although 900 words were recommended by its Ministry of Education). Of the five Dual Assessment Vocabulary Instructor-Evaluator categories listed above, the first three (A-C) may also be given again productively in an active mode. In the later case, students would be asked to write out definitions first in L1 and then in L2. Thirdly they would try to use the word in a sentence. This second assessment could be given either in written form or else by means of an informal individual interview. Students at three colleges in Kyushu were asked to rank their word knowledge using this Dual Assessment

Vocabulary Instructor-Evaluator on paper. It can be also be used as a pre-reading activity to raise both teacher and student awareness of upcoming potential word problems in any text. Five words are included in the DAVIE Samples, as shown below, in Tables 1 & 2, which show both receptive input and productive output samples of using the DAVIE VKS designed and tested at three Japanese colleges to assess students' English word knowledge.

4. Methodology:

Testing of this DAVIE was always done following the typical Action Research cycle of pre-test, experimental treatment (in which learners were required to study and use all words in their own original sentences, using any means they chose), and post-test assessment.

4.1. Pre-Reading Self-Report Vocabulary Knowledge Scale Used

Pre-testing of learners' vocabulary knowledge was done using an author-designed "Dual-Assessment Vocabulary Instructor-Evaluator," first given using a receptive input survey format as described here. This Evaluator is given receptively the first week, simply allowing students to rate their own knowledge of these or any other target vocabulary words by simply putting checks under any appropriate categories of word knowledge, from A-E. Next, the "Productive Assessment" Evaluator is usually given separately one week later, it less time-consuming for teachers and easier for students to merely check as a survey first. When giving this second productive assessment usually one week later, for words believed to be known students write in L1 translations and L2 synonyms they think they know under columns A & B, writing complete sentences for C on the back of the test form. A sample assessment of five words is shown in Table 1.

A. Experimental Treatment

All but one class of Engineering and Foreign

Designing and Testing a User-Friendly Vocabulary Knowledge Scale

Table 1: Dual Assessment Vocabulary Evaluator for Japanese Students (Receptive Assessment)

Recognition: Know L1 Japanese Translation	Recognition: Know L2 English Definition	Self-Report: Think I Can Use Word in a Sentence	Unclear: Have Heard, but Not Sure	Unknown: Word; No Idea at all	Word Token or Family	Modified ICU EAP List #
A (60 %)	B (40 %)	C (20%)	D (20 %)	E (20%)		
X	X	X			abandon	1
X	X				abbreviate	2
X			X		abide	3
				X	ability	4
					abnormal	5

Date: / / Circle: T1 Name: 3/5= 60% L1 Translation Ability

Table 2: Dual Assessment Vocabulary Evaluator (Sample of Productive Output Assessment)

Recall: Show L1 Japanese Translation	Recall: Show L2 English Definition	Generative Production: Use this Word in a Sentence	Unclear: Have Heard or Read in this Phrase	Unknown: Word; No Idea at all	Word Token or Family	Modified ICU # EAP List
A (60 %)	B (40%)	C (20%)	D (20 %)	E (20%)		
2 Points	2 Points	4/5 Sentence	1 Point	No Points		
みすてる	leave	I leave home			abandon	1 10 points
はぶく	shorten		X (great ability)		abbreviate	2 5 points
とどまる				X	abide	3 2 points
					ability	4 1 point
6 Points	4 points	5 points	1 point	0 points	abnormal	5 0 points
					Total Score:	16/50 points

Date: / / Circle: T2 Name: 16/50=32% Productive Output Assessment

Table 3: DAVIE Comparing Student's Self-Reports with Productive Output Assessment

A: Dual Assessment Vocabulary Evaluator Sample of Japanese Student's Self-Reports

Recognition: Know L1 Japanese Translation	Recognition: Know L2 English Definition	Self-Report: Think I Can Use Word in a Sentence	Unclear: Have Heard, but Not Sure	Unknown: Word; No Idea at all	102 Items Important Adjectives (or Verbs)	1-20 Short Survey
A (51%)	B (25%)	C (36%)	D (23%)	E (22%)		
53/102	27/102	38/102	24/102	22/102		

Date: / / Circle: Pre-Test Category Percentages T1 Name:

B: Dual Assessment Vocabulary Evaluator (Sample of Productive Output Assessment)

Recall: Show L1 Japanese Translation	Recall: Show L2 English Definition	Generative Production: Use this Word in a Sentence	Unclear: Have Heard, but Not Sure	Unknown Word; No Idea at all	100 Items Important Adjectives (or Verbs)	1-20 Short Survey
A (58%)	B (28 %)	C (39%)	D (18%)	E (19%)		
60/102 +7 words +7%	29/102 +2 words +3%	40/102 2 words +3%	19/102 -5 words -5%	20/102 -2 words -3%		

+7← +2← +2← A←-5 A←-1 (Gain or Loss per Category Shown)

Date: 12/1 /02: 39 % Productive Use Output Assessment T2 Name: H-san

Table 4: Average Dual Assessment Vocabulary Instructor-Evaluator Results

A. For the EAP Summary List (Adv.=Advanced; Inter.=Intermediate; Thot=thought)

Sample/ School Where VKS Tested	A Japan. Trans. Thot Known	B Thot Known English Definition	C Think They Can Use Word	D Unsure of Definition	E Word Not Known	F Knew Japanese Translation; % reliable	G Knew English Definition; % reliable	H Used Correctly in Sentences; % reliable	I Gates Vocab. Level (Native Norms)
Eng 1 Pre- Advd.	50.15%	25%	31.23%	26.23%	22.5%	45% accurate 0.897 reliable	15% accurate 0.60 reliable	27.5% accurate (.88 reliable)	4.7 (n=18)
Upper Inter. English	23.5%	20.8%	27%	31.6%	58.5%	23% Valid: 0.98	20% Valid: 0.96	25% Valid: 0.93	3.9 (n=11)
Lower Interm. Eng 2	25%	3%	14.36%	17%	58%	24% 0.96 reliable rate	3% 1.00 reliable rate	11% Correct 0.77 reliable	3.3 (n=17)
Lower Interm. Electric	8%	4.4%	2.5%	19%	76%	6% 0.75 reliable	4% 0.91 reliable	2.5%; 100% reliable	2.5 (n=8)
Aves.						89.97% accurate	86.75% accurate	89.5% accurate	3.6 (n=54)

(Summarized results for the first 100 word families on Mizoguchi et al.'s Recommended EAP List, 1992.)

B. Foreign Students' Average Knowledge of 102 Adjectives (JACET English Usage SIG, 1995)

	Nationality	L1 Translation	L2 Definition	Use in L2 Sentence	Unclear	Unknown	Student
		A	B	C	D	E	
102 Adjs	J	69.6	64.7	78.4	70.59	0	E JA
	K	NA					Wi K
	C	35	3.7	3.7	18	48	Z C
	J	60	57	57	32	14	M J
	J	30	9.2	0	24	33	A J
	CM	NA					L Y CM
	J	NA					M J
	C	20	12	22	21	35	SAi C
Averages		42.92	29.32	32.22	33.118	26	

(J=Japanese, K-Korean, C=Chinese, CM=Chinese-Mongolian)

Student English class students were given the treatment of doing a "Pushed Output" assignment of using all target terms in their own sentences productively. Control group classes were not given any such assignment, but were merely assessed to see what difference such guided instruction would have upon their subsequent lexical development.

B. Post-tests were given using the Productive Output format shown in Table 2.

For "Productive Output Assessment" of words believed to be known, students write information they think they know under columns A & B, writing phrases or sentences for C on the

back. A simpler, non-weighted rating is just to use 10, 20, or even 50-100 words at the beginning of instruction to *determine average percentages in each category of both receptive and productive knowledge* of certain target language vocabulary. Teachers or researchers can easily compare these rates. Category C is explained as:

- 1) Add another English word that you often think about/connect, read or hear when you see this word being used (common co-occurring collocations or correct phrases= 2/3 points).
- 2) Make a sentence using this word correctly to show you know its meaning (clear

and correct sentences gain 4-5 points, depending on semantics and syntax).

A somewhat more lengthy rating procedure is to use the above weighted scale, which also generates a percentage, but knowledge of L2 definitions and ability to produce L2 sentences are both more heavily weighted, so that higher scores will reflect a higher degree of knowledge in these important areas of language usability. This second weighted rating procedure is done as follows. Each word or sentence is worth 1-10 points as shown in the chart above (Table 2). That is, a word that appears to be familiar is worth 1 point, whereas if the student can actually give a) Japanese definition (2 points), b) English definition (3 points), or c) Use the word in a sentence with clear meaning, these are worth 4 points, or 5 points if used with correct grammar, that is, in sentences showing both correct semantic meaning and also a correct understanding of the word's grammatical form and functions. Thus, a "Perfect Productive Score" would be 10 words x 10 points each=100%.

However, the simpler way of scoring for busy bilingual teachers is to simply compare all of a student's "Objective Output" responses with their "Subjective Self-Report" responses, evaluating percentages of target words thought to be known against those whose definitions and sentence usages are in fact given correctly. This type of comparison is made in Table 3. It is much less complicated than Paribakht and Wesche's (1997, p. 181) VKS survey procedure, which according to Read's (2000, p. 132) critique "consists in effect of two scales: one for eliciting responses from the test takers and one for scoring the responses." While in his words (Read, 2000, p. 136), "it is doubtful whether learners' developing knowledge of second language words can be meaningfully represented by a single linear scale," use of a more clear-cut instrument such as this Evaluator can help to show learners and instructors both quantitative and qualitative change, since students are asked to make a final

chart for each, showing total words that have been learned or changed categories. In this way learners can show and teachers quickly observe by means of simple arrows as well as on post-test composite forms how many words have moved and into which categories of word knowledge, as shown here in the Sample presented in Table 3.

At first this Evaluator is simply given as a subjective checklist of *receptive word knowledge*, in which students simply put a check under whichever column of word knowledge they believe applies. Later for further testing and instruction, words checked by each student as completely unknown may be highlighted and focused on in instruction or used for research purposes. This Evaluator was first used to determine to what degree the first 100 word families on the *Recommended EAP List* (Mizoguchi, et al., 1992) were known by each student. Later during ten-minute periods they could search to find definitions of words they had indicated were unknown to them individually, by using various computerized tools and methods, which were compared to using a traditional bilingual book dictionary. Results of that computerized bilingual dictionary study are beyond the scope of this report. For details comparing use of each of these computerized versus traditional bilingual book dictionaries, see author's other studies.

Briefly stated, students were taught how to use and compare various types of dictionaries--both bilingual book and computerized dictionaries for ten minutes each, in order to better develop their vocabulary acquisition skills and strategies. Target words were first determined to be unknown using this Evaluator.

Secondly, this Evaluator was used to compare eighty Japanese college students' learning of important verbs and adjectives. The Japan Association of College English Teachers (1995 & 2000) English Usage Special Interest Group did two studies, one called *A Study of Synonymous Verbs in English*, and the other *A Study of Synonymous Adjectives in English*, from which words were drawn. General overall lists of 102

important common adjectives and 109 basic verbs from the indices of these works were used for longer surveys, with shorter twenty-item surveys taken from the 21 main adjectives and 23 verbs investigated comparatively and bilingually within the contexts of these JACET English Usage SIG studies. In this way, the Dual Access Vocabulary Instructor-Evaluator was used and assessed in a variety of vocabulary learning contexts for EAP words, for essential adjectives and verbs, and in both shorter (20-item) and longer (100-item) contexts.

Thirdly, Paribakht and Wesche's original VKS was compared with this proposed DAVIE VKS in one class of thirty-eight 2004-5 graduate engineering students, to better assess the relative strengths and weaknesses of both instruments.

In the computerized studies of EAP vocabulary acquisition using this instrument (See Author, 2002a, 2002b, and 2002c), students' type of book dictionary or computerized translation device was carefully monitored and guided by the teacher. In studies of adjective and verb acquisition done since 2002, on the other hand, students' type of dictionary access was not controlled, but rather left up to learners' autonomy or individual choice. These newer studies have focused on examining the usefulness of this Evaluator for both *pre-instructional assessment*, as well as *post-instructional reassessment* of each student's quality and quantity of vocabulary learning, to help measure change in each of its five areas.

4.2. Participants

This paper simply seeks to present the design, testing and results of this new Dual Assessment Vocabulary Evaluator as it was used over three years with these participants: 1) At three colleges (n=54; 50 Japanese and four Chinese) in Kyushu from October to December, 2000. 2) Secondly, by eighty (n=80; 69 Japanese, 3 Koreans, and 8 Chinese) other students at those same schools from April, 2002 to February, 2003. Finally, 3) Thirdly, by thirty-eight (n=38) Japanese graduate students of engineering from September 2003-March 2004,

only one of whom was female. This third group was randomly divided and half used the author's DAVIE VKS to assess their knowledge of twenty words met in assigned online readings, whereas the other half used Paribakht and Wesche's (1993) original VKS. The three language group comparison used 15 students at one women's junior college of economics, of whom 3 were Koreans, 4 Japanese, and 8 Chinese.

5. Results

Reported as briefly as possible, the following chart summarizes students' *perceived* versus *actual word knowledge* of these initial 100 EAP vocabulary words, giving a degree of self-report to actual score accuracy as a percentage of these two measures. That is, students' average subjective self-rating percentages were divided by the percent of their objective actual word knowledge to determine the overall accuracy for each category of this Dual Assessment Vocabulary Evaluator.

The Posttest given to this class with three language groups in it is most enlightening in making linguistic and cultural background comparisons. It showed us average productive vocabulary learning rates for these three different Kanji-background Language Groups, after instruction in "20 Essential Adjectives and 20 Essential Verbs" (chosen from JACET Studies). All three groups started at very comparable positions of word knowledge, knowing mother tongue translations or meanings for only 10-12% of the target language vocabulary (TLV), according to their own self-reports prior to instruction.

Clearly this DAVIE VKS proved to be extremely helpful in correctly and quickly assessing the vocabulary knowledge and learning of these students from all three Kanji-background language groups. Linguistic comparisons will now be given to show that use of this Vocabulary Evaluator for measuring both "Reported Receptive Vocabulary"

and Objective Productive Vocabulary” knowledge is both simple and straightforward, and that anyone may easily replicate such results with similar students or use this instrument easily and practically with other language pairs.

The following table compares the vocabulary learning of three groups of students all from countries where they use Chinese or Kanji characters in their reading and writing systems. Pre-Test DAVIE “Receptive Input Word Recognition” scores are compared with students’ Post Test DAVIE “Productive Output Word Recall and Use” to assess their overall learning of 40 common words, specifically 20 Verbs and 20 Adjectives from JACET’s English Usage SIG (1995, 2000 research publications). These were their results after they were given these words to study and practice using:

C. For Students’ Average Knowledge of 20 Verbs & 20 Adjectives

1) Pre-Test Self-Reports versus Posttest Average Learning Rates:

I. Learning of 20 Common Adjectives:

- A) Koreans—Were able to use target language vocabulary (TLV) adjectives in sentences correctly only about two-fifths or 40.33% of the time (partial credit given for correct collocations, such as “deep down in my heart”, basic language,” “get fat,” “low tide”, or a “clean room”).
- B) Chinese—Were able to use target language vocabulary adjectives in sentences correctly two-thirds, or 66.88% of the time.
- C) Japanese—could use them properly in English sentences 64% of the time.

II. Learning of 20 Common Verbs:

- A. Koreans—Were able to use target language verbs in sentences correctly 43.67% of the time (partial credit given for correct collocations, such as “postpone until the following day,” or protect a person from”). Often they wrote nouns instead of verbs. They made several errors

due to phonological and morphological confusion of similar words, reading pull as full, and reach as search, etc. Due to the phonology of Korean it seems difficult for them to distinguish between similar sounds, such as f from p, p from b, t from d, etc.

- B. Chinese—Were able to use target language verbs in sentences slightly under two-thirds, or 60.5% of the time.
- C. Japanese—Could use verbs in sentences correctly 62.75% of the time, close to two-thirds of the time, and only about 1% less than they could use common adjectives.
- D. Combined Posttest Individual and Language Group Averages:

When one considers that these students were mostly foreign students studying both English and Japanese abroad for the first time as Freshmen, and that both initial surveys taken using the DAVIE Vocabulary Knowledge Scale and objective corrections thereof showed that they could not demonstrate an ability to use virtually any (in fact an average of only 1/3rd of 1% or .36%) of these most common English adjectives and verbs at the start of instruction, their learning rates were quite good. Table 5 below compares these rates, which may be summarized as follows:

- 1) Adjective Learning Rate for all students—averaged 53.28%.
- 2) Verb Learning Rate for all students average 52.06%.
- 3) Total Average for all students on the combined Posttest was 53.42%

B. Chinese students, on the other hand, who began knowing only 1% of these words with an English definition, and no ability to use any target word in a sentence, after instruction could use 1) 66.88% of these 20 basic adjectives correctly in sentences, 2) 60.5% of these basic verbs correctly, and 3) a combined 63.81%, or slightly less than two-thirds of such basic words correctly in whole sentences. (Chinese Average Vocabulary Level=3.08; n=8).

- C. Japanese students, originally could use

Table 5: Chinese, Japanese and Korean College Students' Learning Rates for Adjectives and Verbs

Foreign Students:	L1 Translation	L2 Definition	Use in L2 Sentence	Unclear	Unknown			
PreTest on						PostTest%	20Adjs	20Verbs
Adjectives:	A	B	C	D	E	(20 Verbs &		
JA=Japanese-American						20 Adjectives)		
E (JA)	NA					76	76	75
A	10	2	0	6	2	38	33	43
M	15	6	0	1	2	91	92	90
M	10	2	0	3	4	49	55	43
Japanese Ave:	11.67%	3.33%	0%	3.33%	2.67%	63.5%	64%	62.75%
S	14	4	0	5	1	88	90	86
Li	6	0	0	4	2			
Ma	6	0	0	3	8	NA		
T	15	0	0	5	0	67	66	66
S	19	0	0	1	0	81	81	81
J	6	0	0	4	9	6.5	4	10
Li	14	0	0	6	0	84	86	83
W	NA					56	67	46
He						55	62	47
Z						72	79	65
Chinese Ave	12%	1%	0%	4%	3%	63.81%	66.88%	60.5%
Wi	NA					27	28	26
I	12.5	2	2	0	6	51.5	45	50
Ra	8	0	4	1	10	47.5	48	55
Korean Ave:	10.25%	1%	3%	.5%	8%	42%	40.33%	43.67%
Class Aves.	11.5%	1.4%	.36%	3.45%	2.8%	53.42%	53.28%	52.06%

no words in sentences, though they could define about 3% in English. After instruction 1) they could use 64% of these 20 basic adjectives correctly in sentences, 2) 62.75% of these basic verbs correctly, and 3) a combined 63.5%, or could use about two-thirds of these essential words correctly in whole sentences. (Japanese Average Vocabulary Level=3.58; n=4).

D. Combined Averages for All Language Groups: (N=15; Average Vocabulary Level=2.88).

First of all the entire class went from a class average of not being able to define more than 1.5% of these forty words in English, or use any of them at all in sentences to the average student being able to use just over half, or 50.64% of them in sentences correctly. Averages were just about the same for both Chinese and Japanese students, at 63.81% for the Chinese and 63.5% for the Japanese, while these Korean students could use an average of less than half, or 42%

of words studied. Averages for the total class in ability to use these 20 "Common Adjectives" was 53.28%, only slightly better than their ability to use the 20 "Common Verbs," on which they averaged 52.06%. Combined posttest average was 53.42% for both. Validity percentages give simple objective percent of student self-report responses that were correct when examined by inter-test reliability checkers in all four languages (English, Japanese, Chinese and Korean).

Finally, results of comparing Paribakht and Wesche's original VKS with this proposed DAVIE VKS by graduate engineering students, found these relative strengths and weaknesses. The bottom line conclusion is that what students actually know and can do with target words is much more important than what they think they know. In other words, what foreign language learners think a word means, versus feeling that they know what a word means really makes little

practical difference, since less proficient learners often seem to misperceive or misread so-called *false friends or synforms*. So once again what is most important for language teachers is to have a tool by which they can quickly ascertain whether particular learners ACTUALLY DO KNOW a certain target term or not, requiring further explanation, clarification and instruction. This DAVIE instrument appears to fit that need rather well.

These two VKS forms differ greatly, in that the DAVIE first quickly determines if target terms are at least understood conceptually in the mother tongue in its first column A. On the other hand, Paribakht and Wesche's VKS (1993) tends to confound both students as well as these degrees of certitude, which being subjective self-reports are not reliable enough in the final analysis of word knowledge. A common problem for many of these Asian learners, and perhaps for other lower proficiency learners with little natural exposure to the target language, is that the original VKS categories overlap, and are thus harder to understand in terms of the learning task or scoring method. Many learners surveyed using Paribakht and Wesche's VKS in fact failed to realize the need to write down their understood meanings or use any target terms in translations, definitions or sentences, although they often claimed having seen them before. Either they just could not produce this kind of knowledge, or they were confused as to which kind of word knowledge they should demonstrate in writing and in what way. This was in spite of reading and

hearing both surveys explained clearly in their mother tongue before being given the surveys, as in Table 6.

Paribakht & Wesche's Word Knowledge Scale scores any sentences given with 5 points for correct meaning and grammar, and 4 points if meaning is a target word's meaning is properly communicated but not in a grammatically correct manner. Whereas this is DAVIE's topic C, it's the last item in the original VKS. Although the original VKS begins by asking students to check either:

1) I don't remember having seen this word before, or 2) Though I have seen it, I don't remember its meaning, the DAVIE is much more straightforward and clearly distinguishes between these by asking learners to first give A. Native translation, and then B. English definition of any word claimed known. Comparing subjective memories of words or degrees of certitude about word meaning are far too ambiguous and abstract tasks for most teachers to be able to explain to foreign language learners sufficiently in L1 or L2, as we found from trying to do so. Instead, the DAVIE instrument is straightforward and practical, and being production- or performance-based, it quickly ascertains what individual learners ACTUALLY KNOW AND CAN DO with certain target terms. This is vastly more helpful to teachers who have little extra time to explain or analyze results of complicated rating scales, or compare subjective nuances of word knowledge.

The DAVIE not only ascertains key aspects

Table 6: Comparison of DAVIE and Paribakht & Wesche's Word Knowledge Scales

	A	B	C	D	E
DAVIE	6.2	0.05	0.5	4.8	9.79
DAVIE%	31	0.25	2.5	24	48.95
vs. P&W VKS	7.68	1.84	0.75	6	.93
P&W %	38.4	9.2	3.75	30	4.65
Compare DAVIE	1) E	2) D	3) A/B	4) A/B	5) C 4/5 pts.
P&W Topics:	Unknown	Seen but Unsure	Think Known	Meaning Known	Sentence Use OK

Table 7: Comparing Self-Reports with Objective Assessment of Vocabulary Knowledge

Aspect of Word Knowledge	L1 Translation	L2 Definition	Use in L2 Sentence	Unclear; Know Phrase?	Unknown	
Pre-Test on						Validity %
20 Verbs:	A	B	C	D	E	(Correct SRs)
A	11	0	0	5	4	100
M	17	0	0	2	1	100
	7	3	3	0	2	80
Japanese Ave:	12.67	1.33	1	2.33	2.33	93.33%
SYA	18	18	18	1	1	100
LM	6	0	0	4	2	100
M	6	0	0	11	2	100
T	13	0	9	4	1	90
S	12	13	8	3	6	90
J	5	0	0	11	4	100
LY	16	13	17	0	2	95
Chinese Ave	12	1	0	4	3	96.43
I	10	0	0	4	2	80
Wi	6	0	0	4	1	80
Ki	7	0	0	4	0	85
Korean Ave:	7.67	0	0	4	1	81.67%
Class Aves	11.5	1.4	.36	3.45	2.8	95.5% Ave

(SR= Student Responses; Pseudonyms used)

of word knowledge possessed by individual learners more quickly and clearly, it also gives us additional information. Whereas the original VKS questions 3 & 4 seem both redundant and foggy (asking "I think it means" vs. "I know it means" and then not requiring a particular category of word knowledge, but instead giving learners an option of producing either native translation or L2 definition), DAVIE asks distinctly different questions for topics C & D. For D they are asked to write down a collocation or phrase use of the target term which they have read or heard. This helps teachers right away to see if they know any basic phrases or idioms using that term, providing them with a launching off point for further discussion or instruction. For topic C, learners are asked to use target terms in sentences of their own. These may be scored in the same fashion as Paribakht & Wesche did, giving 5 points for sentences that are both semantically and syntactically correct, 4 points if meaning is but grammar is not. Detailed rating and analysis is not always needed, desirable, or possible given

time constraints, but can provide teachers with further helpful clues as to learners' common error patterns or misunderstandings.

Examples of sentences written that give useful insights to learners' mental lexicon and grammatical misunderstandings are these: 1) "The consequence is clearly." -1 for using an adverb where an adjective belongs. 2) "I have a motivation with work." -1 for wrong preposition in place of "to work." Many false friends also appear.

6. Discussion

When one considers objectively how well this instrument performed to enable better measuring and motivating of basic vocabulary instruction one can quickly see its clear practicality and effectiveness. For an instrument such as this Vocabulary Instructor-Evaluator to be readily understood by foreign language students from three different backgrounds is challenging

Table 8: Dual Assessment Vocabulary Evaluator: Total Averages at Three Japanese Colleges

A Japanese Translation Thought Known	B English Definition Thought Known	C Think They Can Use Word	D unsure of Definition	E Word Not Known at all	F Knew Japanese Translation (A)	G Knew English Definition (B)	H Used Correctly in Whole Sentences (C)
26.7%	13.3%	16.1%	23.5%	53.7%	24.5% 0.92 accurate	10.55% 0.79 accurate	14.5%; 0.90 accurate

enough, all of them at a low intermediate level of English, with an average of just grade 2.88 level vocabulary. For the instrument to have worked so well that such learners could easily note their level of vocabulary knowledge by its five simple categories, and then study accordingly, with learning rates averaging from 40-60% is positive proof of its effectiveness.

This multi-language background class was an ideal setting for checking on the usefulness and understandability of this new DAVIE VKS, and its results were also compared when used with several other fully Japanese college classes. Table 8 shows total averages and overall accuracy figures for the Evaluator described above, when all Japanese college student scores in 2002 were combined.

In other words, usually the numbers of words thought to be known on this Evaluator were reported with very high accuracy (rates of 79 to 92%), regardless of the level of students' English proficiency. These results show that it is a very helpful and quite reliable tool for assessing and distinguishing between these two types of vocabulary knowledge. Specifically, Japanese definitions actually known were only 2.2 less than those thought to be known (24.5% rather than 26.7%). English definitions actually known were only 2.75% less than those thought to be known (10.55% rather than 13.3%). Even clearer was the fact that whereas students thought they could use in sentences an average of 16.1% of the first 100 modified EAP List words, 14.5% of those words were used correctly in terms of meaning. Grammar was not a consideration. There was only an average 1.6% gap between words reported as known versus Japanese students' actual ability on this Evaluator. With such small gaps of only 1.6

to only 2.75% between reported and actual scores, this Evaluator seems to have strong support based on both its understandability to students and its high degree of correspondence between students' *reported versus actual word knowledge*.

While student self-reports are not always accurate, the information they supply is exactly what teachers need to gain a clearer perspective into each individual's thinking and mental lexicon, in both L1 and L2 if such teachers are bi- or tri-lingual. Using the results of the Verb Quiz above, for example, one Japanese student thought she knew 10/20 (half) of the target words in Japanese, her L1. Her self-report survey, however, revealed that while seven were correct, she had 3 misreadings. Once again these were all *synforms* or *false friends*, namely she misread the verb *bear* as a noun, the animal *bear*; *postpone* as *post office*, and *reach* as *search*. Quickly being able to know this kind of detailed information for each individual learner, even for those of various language backgrounds, is most invaluable for a teacher. At times, of course, teachers may need to enlist the help of a bilingual assistant or ask one of the better students to help check native language guesses which they cannot read for themselves. The other two Japanese students reported their vocabulary knowledge with 100% accuracy. Even when student reports are not 100% accurate, they can be checked quickly and easily by any teacher who is bilingual for category A, and by any English teacher for other categories. Inter-rater reliability checks done using bilingual Korean-Japanese-English speakers showed 95-97% agreement in interpretation of student scores.

When comparing objective DAVIE VKS pre- and post-instruction survey results across language groups one can also say that the effects

of requiring “pushed output” (See studies thereof by de la Fuente, 2000; Swain, 1995) via written use of these target terms in original sentences of their own, followed by encouragement of greater generative output via individual study and oral interviews with partners had remarkably good results in obtaining high class average vocabulary learning gains in terms of both retention and increased ability to actively use both targeted verbs and adjectives productively.

To summarize these charts, regardless of these language learners’ English proficiency levels, at several different colleges the use of this Dual Assessment Vocabulary Evaluator seemed to reflect fairly accurately each student’s individual actual L2 word knowledge, as well as each group’s average word knowledge when so calculated. According to expectations, as levels of proficiency declined (as shown in Table 3, Column I), so did the percentage of words thought to be known (A-E), and average rates of active production (F-H) declined as well. In addition, at each level of proficiency, students’ self-assessed *receptive word recognition* levels were fairly close to their actual *active productive vocabulary* abilities. Total averages showed a fairly close correspondence as well, as the following comparison bears out.

When comparing areas A-C, *passive recognition* of EAP words thought to be known by students with *active productive use*, one should look at actual results in areas F-H. Although some may question calling recall of an L1 or L2 definition “productive output,” it is used here as such for two reasons. 1) First of all letters F-H are only given as part of the “Productive Output Assessment” part of the test. 2) Secondly, since Asian students tend to be used to overly passive, teacher-fronted language education, any oral or written response may be seen as actively giving evidence of word knowledge. Here we would use the *recall* versus *recognition* distinction to delineate between *active production* and *receptive knowledge of words heard or read*, which students only show by checking items A-C the first time the test is given.

Results when using this Evaluator as an aid in guiding subsequent vocabulary instruction, use and study may be summarized as follows. The DAVIE proved to be highly effective with language learners from different backgrounds, since the teacher could both assess and focus on addressing very particular TL vocabulary mis-readings at specific points of misunderstanding. In a mixed class of ten foreign students and four Japanese, 11 of Japanese student A’s errors, (*postpone* misunderstood as *post office*, *reach* misunderstood as *search*, and the verb *bear* mistaken for the noun, *bear* the animal) for example, were *synforms* or *false friends*, meaning words which can easily be mistaken for similar looking or sounding words. Student B knew 17/20 verbs, but was unclear about one word which she guessed incorrectly, and she did not try to guess two others. Student C also mis-guessed three *false friends* as follows—1) taking *wise* as *wife*, 2) *low* as *law*, and 3) *broad* as *abroad*. The Japanese-American student had a developmental learning disability, so her incomplete surveys were not counted.

In comparing the two Korean students’ responses it was learned that not only is category C’s distinction between semantically correct meaning and grammatically correct form a helpful one, but in terms of various aspects of measuring vocabulary knowledge, another distinction is perhaps even more important to measure with this category. That is the distinction between collocational knowledge shown when a student can use target words correctly in a phrase with common collates, versus the ability to use the term correctly in a whole sentence, requiring and measuring more syntactical ability rather than just simply semantic knowledge.

Specific examples may be given from Korean students’ answers to illustrate this. One student misread the adjective *careful* as *care for*, giving both the Korean term for that and also adding an English related term, *care of* someone. For related words it seems one should therefore assign half credit (.5 points), in order to show another aspect

of lexical knowledge was demonstrated for a particular category. Since use of this DAVIE continuum can also help to show what particular type of lexical misreading or misunderstanding took place, it is most helpful for more focused teaching and correction. The more specifically focused and precisely diagnosing a test is, the more useful it is in helping to provide better description and prescription of corrective remedial instruction.

One Korean student showed knowledge of the noun form of wise (wisdom), even though she could not provide an English definition of the term. She also added two antonyms for beautiful (plain), and clean (dirty), showing the instructor that while she did not have knowledge of synonyms, she did know the antonyms for these terms. Thus, students should be encouraged to write any English knowledge they may have of these target terms, even if they cannot write either an English definition (B) or a full sentence (C), which most of them cannot or do not even attempt. So for Category C students should be encouraged to write just a phrase using the target word, even if they can't use it in a sentence, since this can helpfully reveal their knowledge of collocations or mental confusions to a teacher. For example, for the word broad, a Korean student did not show knowledge of the expected answer *wide*, but rather knew the meaning of *generous or broad-minded*, listing it also as having a *deep heart*. Another Korean student could write four English sentences, but for four other words only could write four phrases showing correct knowledge of their common collocations (*electric dictionary, basic language, low bench, and high building*).

A Japanese student read the verb *bear* as *beer*, and used an archaic though formally correct definition for *appropriate*, the word *meat*, as in fitting. He also wrote a sentence saying "We shaked hands," rather than *shook*. Using such an easily administered VKS survey can help any teacher to quickly zero in on each student's individual word errors, whether those mis-readings are due to misunderstandings in

grammar, meaning, spelling, pronunciation, etc. Such insights of what foreign language learners are thinking are most useful to teachers, and even those who are not bilingual (or trilingual) can gain a much better perspective into each learner's L2 mental lexicon by using this DAVIE.

In the graduate school comparison of this DAVIE VKS with Paribakht & Wesche's original (1993) VKS, a few other observations may be made from Table 6: Comparison of DAVIE and Paribakht & Wesche's Word Knowledge Scales. Whereas one can clearly see that 31% of these target words were translated corrected on the DAVIE survey (double-checked by a native), one cannot determine whether learners chose to give Japanese translations or English definitions easily, as these are combined and therefore confounded in the original VKS questions 3 & 4. Furthermore, from the DAVIE one can determine just how weak these learners are in terms of inability to define target terms using the L2, learners only being successful in .25% of cases. If using the original survey, it would be anyone's guess, as its questions 3 & 4 not only confuse students by asking if they "think they know" versus "know" a meaning (a fuzzy/foggy distinction to many of them it appears), but these questions also fail to disambiguate between L1 versus L2 translation versus defining ability, which are vastly different in anyone's mind.

The DAVIE also can help us to zero in on individual or class knowledge of collocations, here showing us that they knew an average of about 4.8 out of 20 terms, or 24% of them. Finally, in terms of ability to actually use terms correctly in original sentences, the DAVIE showed these learners could do so for only 2.5% of words. About 5% (4.65) or only .93 out of 20 words could be used by the other group, again showing great productive weakness among these Japanese graduate learners, despite having had over 7 years of prior English instruction.

DAVIE seems to give much clearer language use tasks to students, who need more obvious categories, given their limited language

proficiency. This Evaluator also clearly tells students what to write to demonstrate actual vocabulary knowledge versus only perceived knowledge. The bottom line difference between these two kinds of VKSs is that between an invisible "I think I know" versus much more visible, clearly demarcated demonstration of particular categories of word knowledge: "Here, I'll SHOW you what this word means to me!" Clearer parameters of the DAVIE appear to result in faster student learning as well as teacher use of this assessment instrument. Students quickly learn how to total their own scores. They can hardly figure out what to write versus check on the original VKS, our results clearly show. The latter may have been designed as a teacher's evaluation tool, rather than for student's own in-class use.

7. Conclusions and Pedagogical Applications

Naturally, self-report surveys do not have the same validity which objective tests may have. Nevertheless, this Dual Assessment Vocabulary Instructor-Evaluator seems to have an acceptable margin of error when one compares students' self-reports with results of actual objective scoring of their L1/L2 translations, definitions and sentences. These results ranged from 79-92% accuracy, depending on the classes and proficiency levels of students being tested. Surveys can never claim 100% accuracy in any field, as respondents could lie or be mistaken in their self-reports. For this reason, careful, detailed individual comparisons were made of *reported* versus *actual* L2 vocabulary knowledge, from which class averages were computed (shown in Tables 3 & 4), as well as a comparison made of both VKS surveys used with forty learners.

By being more easily read and understood by students, a user-friendly VKS like this Dual Assessment Vocabulary Instructor-Evaluator can help both teachers and students to more clearly distinguish between types of vocabulary

knowledge in ways that mere percentages or word counts cannot easily do. This is because arrows on the Evaluator's post-test chart can easily show us from which category of lexical knowledge words have moved (see Table 2 for an example), and into which new level of understanding. Other VKS generally are not designed to give students such a clear and composite picture, quickly portraying individual changes in lexical knowledge. These changes can also easily be averaged to see a class profile as a whole, as was done for these classes using *Excel*. One can also determine what amount of growth or attrition has taken place in which particular areas of lexical knowledge, either in percentage terms or else by number of words gained or lost per category, in order to help language learners better focus their study on most essential areas of weakness that are thereby revealed.

Since the DAVIE's design is user-based, for in-class use in both vocabulary instruction as well as evaluation, it works more practically and efficiently than many longer tests with more time-consuming scoring procedures, as it involves learners themselves in totaling their own assessments, which teachers can quickly confirm or correct (Author, 2002c, 2005b). Categories of vocabulary knowledge are quite clear, so even if students write in wrong columns moving scores for correct assessment is quick and easy. The clear, simple column headings marking each of the aspects of word knowledge on the DAVIE make it much easier for learners as well as teachers to understand. Ambiguous categories in the original VKS tend to create confusion on the part of many lower level language learners. Thus this new type of VKS should be a welcome addition to the arsenal of any language teacher, especially those wanting to get a "quick read" of either individual or overall class vocabulary knowledge about any particular list of target words before trying to teach them.

References

- Alderson, J.C. (2000). *Assessing Reading*. Cambridge: Cambridge University Press.
- Brown, J.D., & Yamashita, S. (1995). English language entrance examinations at Japanese universities: What do we know about them? *JALT Journal*, Vol. 17, No. 1, May, 7-30.
- Chapelle, C.A. (1994). Are C-tests valid measures for L2 vocabulary research? *Second Language Research* 10, 157-187.
- de la Fuente, M.J. (2002). Negotiation and oral acquisition of L2 vocabulary: The roles of input and output in the receptive and productive acquisition of words. *Studies in Second Language Acquisition*, Vol. 24, No. 1, March.
- Guest, M. (2000). Which words? A comparison of teacher and learner choices for lexical study. *JALT Journal*, Vol. 22, No. 1, 165-182.
- Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *Canadian Modern Language Review*, 61, 355-382.
- Horst, M., & Meara, P. (1999). Test of a model for predicting second language lexical growth through reading. *Canadian Modern Language Review*, 56(2), 308-328.
- JACET. (1995). *A Study of Synonymous Adjectives in English*. Japan Association of College English Teachers, English Usage Special Interest Group (Gohou Kenkyukai). Tokyo.
- JACET. (2000). *A Study of Synonymous Verbs in English*. Japan Association of College English Teachers, English Usage Special Interest Group (Gohou Kenkyukai). Tokyo.
- Kinoshita, C.J. (2003). Integrating language learning strategy instruction into ESL/EFL lessons. Online at <http://itesl.org/Techniques/Kinoshita-Strategy.html>.
- Laufer, B. (1997). The lexical plight in second language reading. In J. Coady & T. Huckin, (Eds.), *Second Language Vocabulary Acquisition*. Chapter 2, pp. 20-34. Cambridge: Cambridge University Press.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16, 307-322.
- Lessard-Clouston, M. (2000, March). Students' approaches to technical vocabulary learning in an academic context: Relating strategies and success. Paper presented at the Annual Conference of American Association for Applied Linguistics. Vancouver, B. C., Canada.
- Loucky, J.P. (1994). Teaching and testing English reading skills of Japanese college students. *KASELE Kiyō* 22: 29-34.
- Loucky, J.P. (1996). Developing and testing vocabulary training methods and materials for Japanese college students studying English as a foreign language. Ed.D. Dissertation on file with Pensacola Christian College, Pensacola, FL. ERIC Center for Applied Linguistics via fax to (202) 429- 9292; or from UMI Dissertation Services, 30 No. Zeeb Rd., PO Box 1346, Ann Arbor, MI 48106-1346. Online access free from ERIC at: http://www.eric.ed.gov/ERICWebPortal/Home.portal?_nfpb=true&ERICExtSearch_SearchValue_0=JOHN+PAUL+Author+japanese&searchtype=keyword&ERICExtSearch_SearchType_0=kw&_pageLabel=RecordDetails&objectId=0900000b8013f95d&accession=ED406844&nfls=false%20%20%20%20
- Loucky, J.P. (1997). Maximizing vocabulary acquisition: Recommendations for improving English vocabulary learning for foreign language learners. *KASELE Kiyō* 25: 101-111.
- Loucky, J.P. (2002a). Assessing the potential of computerized bilingual dictionaries for enhancing English vocabulary learning. In Lewis, P. N. D. (Ed.), *The changing face of CALL: A Japanese perspective*, pp. 123-137. Lisse: Swets & Zeitlinger.
- Loucky, J.P. (2002b). Comparing translation software and OCR reading pens. On Swanson, M., McMurray, D., & Lane, K. (Eds.), *Pan-Asian Conference 3 at 27th International*

- Conference of JALT, National Conference Proceedings CD*. Kitakyushu, JAPAN. Pp. 745-755.
- Loucky, J.P. (2002c). Improving access to target vocabulary using computerized bilingual dictionaries. *ReCALL* 14 (2), pp. 293-312.
- Loucky, J.P. (2002d). Testing vocabulary levels in Japan. *The Japanese Learner*. Oxford: Oxford University. Part I (pp. 15-21).
- Loucky, J.P. (2003a). Testing vocabulary levels in Japan, Part II. *The Japanese Learner*, No. 29, (March), pp. 15-20. Oxford: Oxford University.
- Loucky, J.P. (2003b). Using computerized bilingual dictionaries to help maximize English vocabulary learning at Japanese colleges. *CALICO Journal*. September Issue.
- Loucky, J.P. (2004). Clarifying and resurrecting the Semantic Field Keyword Approach for rapid lexical acquisition: Employing semantic organization, bilingual computerized glosses and pushed output generation to enhance L2 vocabulary learning. *Seinan JoGakuin Tandai Kiyo*, No. 50.
- Loucky, J.P. (2005a). Finding better ways to systematically maximize lexical acquisition by using a depth of lexical processing taxonomy, CALL resources, and effective strategies. *CALICO Journal*, 22, Volume 23, Issue No. 1 (September 2005).
- Loucky, J.P. (2005b) Making a user-friendly Vocabulary Knowledge Scale for Japanese students: Designing and testing a Dual Assessment Vocabulary Instructor-Evaluator. Kitakyushu: *Orio Aishin Tanki Daigaku Ronshu*, No. 39, 33-86.
- Loucky, J.P. (2005c). Systematically employing a trinity of user-friendly assessment tools. *JLTA Journal*, No. 7, 66-91.
- Loucky, J.P. (2005d). Surveying Japanese students' use of electronic dictionaries. *Seinan Women's University Research Bulletin*, No. 9, 187-205.
- Meara, P. (1995). The importance of an early emphasis on L2 vocabulary. *The Language Teacher*, Vol. 19, No.2. Tokyo: JALT.
- Mizoguchi, S., Sano, M., Shiina, K., Thrasher, R., & Yoshioka, M. (1992). A proposal for the establishment of an EAP list and an analysis of its appropriateness. *JACET Bulletin* 23: 77-96.
- Nation, I.S.P. (1983). *Learning and teaching vocabulary*. NZ: Wellington, Victoria University.
- Nation, I.S.P. (1990). *Teaching and Learning Vocabulary*. New York: Newbury House.
- Nation, P. (1994) *New Ways in Teaching Vocabulary*. Alexandria, VA: TESOL.
- Nation, P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, P., & Newton, J. (1997). Teaching vocabulary. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition*. Cambridge: Cambridge University Press.
- Paribakht, T. & Wesche, M. (1993). The relationship between reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal*, 11 (1), 9-29.
- Paribakht, T. & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second Language vocabulary acquisition. In J. Coady & T. Huckin, (Eds.), *Second Language Vocabulary Acquisition*, pp. 174-200. Cambridge: Cambridge University Press.
- Read, J. (1997). Vocabulary and testing. In N. Schmitt and M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy*, pp. 303-320. Cambridge: Cambridge University Press.
- Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Rosszell, R. (2007). Combining extensive reading and intensive vocabulary study in a Japanese university. Unpublished Doctoral dissertation, Temple University Japan, Tokyo.
- Schmitt, N. (1997). Vocabulary learning strategies. In Schmitt, N. & McCarthy, M., *Teaching Vocabulary*, (2.6) pp. 199-227. Oxford:

- Oxford University Press.
- Tsudajuku (1992). Tsudajuku Daigaku Genko Bunka Kenkyu Jo Dokkai Kenkyu Group (Reading Comprehension Study Group of the Institute for Research in Language and Culture at Tsuda College [Tsudajuku]. *Gakushusya chushin no dokkai shido*) [Learner-centered reading instruction]. Tokyo: ACT Taishukan.
- Waring, R. (2000). The 'State Rating Task'--An alternative method of assessing receptive and productive vocabulary. *Immaculata occasional papers of Notre Dame Seishin University* 1: 53-68.
- Waring, R. (2002). Scales of vocabulary knowledge in second language vocabulary assessment. *Kiyo, Notre Dame Seishin University: Studies in Foreign Languages and Literature*, 26 (1): 40-54.
- Wesche, M. and T.S. Paribakht. (1996). Assessing vocabulary knowledge: Depth vs. breadth. *Canadian Modern Language Review*, 53 (1): 13-40.
- Yamato, R. (1997). *Eiken 2-3 kyu reberu gakushusya no dokkai sutorategi ni taisuru ninsiki to siyo no sai* [Difference between awareness and real use of reading strategies among STEP 2nd to 3rd grade learners]. *STEP Bulletin*, 9. 60-71.
- Zimmerman, C. (1997). Do reading and interactive vocabulary instruction make a difference? An empirical study. *TESOL Quarterly*, 31 (1): 121-40.

Appendix A: Author's Dual Assessment Vocabulary Instructor-Evaluator
 (Blank Form may be used to assess any vocabulary word or list.)

A. Simplified Form for Self-Reporting on Receptive Recognition of Input:
 (Use to assess Recognition Knowledge or Recognition Memory)

Recognition: Know L1 Japanese Translation	Recognition: Know L2 English Definition	Self-Report: Think I Can Use Word in a Sentence	Unclear: Have Heard, or Read but Unsure; Can Recall this Phrase:	Unknown Word: No Idea at all	100 Items Important Adjectives (or Verbs)	1-20 Short Survey
A.	B	C	D	E	Word Form	Word #

B. Simplified Form for Objective Assessment of Language Output:
 (Use to assess Recall Knowledge or Recall Memory)*

Recall: Give L1 Japanese Translation	Recall: Give L2 English Definition	Generative Production: Use this Word in a Sentence	Unclear: Have Heard, but Not Sure Try to give in a Phrase	Unknown Word: No Idea at all **	Word Token or Family	Modified ICU # EAP List
A. (%)	B (%)	C (%)	D (%)	E (%)	Word Form	Word #

* This test may be given in either a written or oral form. A written test would assess a learner's word recognition of decontextualized target words first, whereas an oral test would assess both their aural comprehension and then their oral production skills.

** Teachers should aim to provide illustrative sentences for all target words, giving them to students after these two test forms are given. Have them focus on learning missed words in these three steps: 1) by first guessing from the context of these example sentences, 2) writing down what they think those target words mean. Finally, they should be taught to 3) confirm or correct their guesses by asking for correct meanings or looking up any missed words in their dictionaries, learning to use book, portable electronic and online Web dictionaries (See author's www.CALL4All.us Dictionary heading for extensive examples of Online lexicons for 140 languages and thousands of language pairs).

重要な英語動詞や形容詞をアジアの学生に教育するための 使いやすい語彙力スケールの作成と検査

ジョン・P・ラオキ

<抄録>

この研究は新しい語彙を学ぶ時にいくつかの重要な要素がかかわっていることを分かりやすい語彙力スケールを使って調べることを目的とする。この語彙力スケールはアジアの大学生が重要な動詞や形容詞を学ぶ時の道具として用いられた時、語彙力を評価したり、変化を監査するために役立つものである。更に、中国人や日本人、韓国人の大学生たちがよく注意して聞くことができ、評価したり、活性化する方法を学んだり、又重要な英語の動詞や形容詞を習得するために非常に効果的で信頼性があることが証明された。

キーワード：アジアの学生の英語語彙レベル、ボキャブラリーノーリッジスケール (VKS)；語彙力、セカンドランゲージボキャブラリーアクエジション (SLVA)；語彙力の側面を評価する