

Combining Technology and IRT Testing to Build Student Knowledge of High Frequency Vocabulary

Charles Browne

Meiji Gakuin University

browne@ltr.meijigakuin.ac.jp

Dr. Brent Culligan

Aoyama Gakuin Women's Junior College

This article describes a suite of free software programs for cell phones and PCs that have been created to efficiently develop ESL and EFL learner's knowledge of high frequency vocabulary. Until now, this level of efficiency has not been possible due to the variable nature of vocabulary knowledge within a class of students and the lack of diagnostic tools for identifying individual students' known and unknown vocabulary. The programs are capable of accurately and efficiently assessing the learner's English lexical size, identifying which specific high frequency words still need to be taught, and then teaching these important words via a time-intervalled flashcard system and learning games focused on developing automaticity of word knowledge. Although there have been several tests available for making estimates of a learner's vocabulary profile such as Nation's Vocabulary Levels Test (1990) and Meara's Yes/No test (1992), there has been no attempts to identify the specific words a learner knows. Through the application of Item Response Theory to test item responses, we have been able to assign perceived word difficulties to a list of the most common words in English. A computer adaptive test drawing from an item bank of these words quickly and accurately assesses the number of English words known by learners, as well as determines which specific words are known and unknown.

Introduction

Although CALL (Computer Assisted Language Learning) has received much emphasis in recent years (Chapelle, 2003; Ducate & Arnold, 2006; Egbert, 2005; Fotos & Browne, 2004; Hanson-Smith & Rilling, 2006; Levy & Stockwell, 2006; Trinder, 2006), one of the biggest challenges facing schools and universities that wished to utilize technology for language learning had been the high cost involved in setting up and maintaining CALL laboratories. As with many technological innovations, new advances present opportunities to both reduce costs and increase the spread of the technology amongst users. With ownership rate among students of cell phones and MP3 players now far surpassing that of

computers (Browne et al., forthcoming), developing software for these popular devices seems a logical extension.

Combining previous work in the area of CALL (Browne 2004a, 2004b), with research on the importance of developing learner knowledge of high frequency vocabulary words (Nation, 1990, 2001), and the importance of using graded reading materials with low level EFL learners (Browne, 1996, 1998; Day & Bamford, 1998; Nation, 1990, 2001), the authors participated in the development of a variety of online English Language Learning software applications for cell phones and PCs. The suite of programs are capable of accurately and efficiently assessing the learner's English lexical size, identifying which English high frequency words still need to be taught (Culligan, 2008), and then delivering these important words to a system of flashcards and learning games, which focus on developing automaticity of word knowledge through spaced repetition (Ebbinghaus, 1964; Leitner, 1972), extensive graded reading, and listening materials.

After a brief introduction to the rationale for the importance of developing vocabulary size, the software applications, which have been developed to help accomplish this task, will be introduced.

The Movement Towards the Teaching of Vocabulary

The study of second language acquisition (SLA) has seen many swings, from a focus on grammar acquisition to a focus on learning processes. Traditionally, vocabulary learning and instruction were seen as somehow isolated and separate from the mainstream theories of SLA. With the grammar-translation method, and its focus on the syntax of the sentence, it was thought that once the students learned the grammar of the sentences, they would be able to slot in vocabulary and therefore generate language. With the advent of the Audiolingual method, based on habit-formation, vocabulary was again treated in much the same way. Words were taught as replaceable elements within sentence structures that always were the central focus of language learning. Subsequent research has often attempted to account for SLA by looking at grammatical features in such areas as developmental sequences (Cancino, Rosansky, & Schumann, 1978; Pienemann, 1989), the role of input (Loschky, 1994; Shook, 1994; White, Spada, Lightbown, & Ranta, 1991), as well as the role of instruction (Dulay & Burt, 1973; Ellis, 1992; Sharwood Smith, 1981; VanPatten & Cadierno, 1993). From the publication of Corder's seminal paper in 1967 to Larsen-Freeman writing in 1991 on SLA research, the study of grammar and its acquisition has almost become synonymous with SLA.

Concurrent with these developments in SLA, yet somehow apart, certain scholars began to study the needs of the learners from a lexical perspective. Many of the questions they asked, and the results they found are still relevant today. These questions included how many words a student needed to know, how these words should be sequenced, and what the student needed to know about these words. One of the first debates centered on the number of words that a student needed to know. This necessarily led to defining what a word is, and what it means to know a word. While this research primarily focused on first language acquisition, there are obvious implications for SLA as well. The central argument was whether it would be possible to increase a learner's vocabulary by the direct instruction of words and their meaning. If estimates of native speakers' vocabularies were very large, then explicit instruction would not be feasible, and early research seemed to indicate that this was the case. Studies cited in D'Anna, Zechmeister, and Hall (1991) suggested a recognition vocabulary of 155,736 – words (Seashore &

Eckerson) and over 200,000 words (Hartman) but both studies suffered from methodological problems in defining what a word is. Nagy and Anderson (1984) used six semantic categories to organize lexis from a corpus of high school English and found that students were exposed to 45,000 base words and 88,500 word families. They suggested that teaching children “words one by one, ten by ten, or even hundred by hundred would appear to be an exercise in futility” (p. 328), and that teachers should concentrate on teaching skills and strategies for independent word learning. Later research by Goulden, Nation, and Read (1990) questioned whether native speakers actually knew these words. By designing tests based on the frequencies of the words, the researchers determined that native speakers’ vocabulary averages 17,200 words. This number suggests that the learning burden is not as insurmountable as previously suggested. Other research by D’Anna, Zechmeister, and Hall (1991) found a similar result of 16,785 words.

Vocabulary Thresholds for Second Language Learners

According to Brown (1995) an essential component of the development of any pedagogical program is a needs analysis. Before designing and presenting materials, it is imperative to gather “information to find out how much the students already know and what they still need to learn” (p.35). The first piece of the puzzle is to determine what they need to know. In vocabulary instruction, this is a subset of the set of the words students may encounter during their use of the target language. This information can be obtained through the frequency analysis of corpora of English text (Carroll, 1971; Leech, Payson, & Wilson, 2001; Nation, 1990, 2001).

Although there are more than 250,000 word families in the Oxford English Dictionary, which is considered to be the largest dictionary of English in the world, research in corpus linguistics has shown that a very small number of these words are actually used in daily life. In an excellent overview of vocabulary research to date, Nation (2001), found that the 2000 most frequent words of English cover approximately 81-85% of words that appear in general English texts, and that the top 5000 words covers approximately 95% of such texts.

How many words do second language learners need to know? Several researchers have discovered important vocabulary “thresholds” beyond which, second language learners are able to function more successfully and independently. Laufer (1992) compared vocabulary size and reading comprehension scores and found that a recognition vocabulary of at least 3,000 words, which offers approximately 90% coverage of texts, was the minimum threshold for being able to read unsimplified texts (i.e. where there were more readers than non-readers of the text). Hirsch and Nation (1992) found that 95% coverage level (5000 words) represents another important threshold, and that once this vocabulary size was reached, learners were able to read and comprehend texts without the help of a teacher or dictionaries. Unfortunately, EFL learners in most countries do not have nearly this vocabulary size. In Japan, for example, studies by Shillaw (1995), and Barrow, Nakanishi, and Ishino (1999) found that after between 800 and 1200 hours of instruction, Japanese university students had an average vocabulary size of between 1700 and 2300 words, far short of the amount they need to be independent readers and speakers of English.

Assessing a Learners' Vocabulary Ability: The Promise of Computer Adaptive Tests

In order to help students to learn the words they need to learn, the second step then, should be a diagnostic one – identifying each students' vocabulary size, and more importantly, the specific words each of them already knows. Unfortunately, until very recently, the only way to measure a learner's vocabulary size was either to have them check off all the words they knew in a dictionary, or to make a rough extrapolation from random samplings of different frequency bands. The most widely used of such vocabulary size tests is Nation's (1990) Level's Test. Though the Levels Test has proved to be useful as a research tool by profiling the lexical ability of a student, it wasn't designed to identify which specific high frequency words were known or unknown, meaning that test results could not directly inform classroom pedagogy. What was needed was a procedure to assess the perceived difficulty of each word, so that the decision to teach a word could be based on its individual features.

The only statistical procedure sufficiently rigorous to accomplish this type of assessment is Item Response Theory (IRT). IRT posits the hypothesis that the probability of getting a correct answer to an item depends on the difficulty of the item and the ability of the student. IRT allows us to be able to measure a test-taker's ability by assessing his or her responses to questions (items) of known difficulty. IRT-based tests are uniquely suited for the creation of Computer Adaptive Tests (CAT) and have recently been employed by large testing companies such as ETS who use it for the online version of TOEFL.

In a CAT, each item presented to the test-taker is selected to provide the maximum amount of information possible toward establishing an estimate of the test-taker's ability. Unlike conventional tests where a reliability index is calculated post hoc, the CAT draws on items until a desired degree of accuracy is obtained. The amount of time necessary to take the test is variable because the process depends on the responses of the test taker. However, because each item is selected to maximize the information and minimize the error based on how an individuals responses, these tests are always more efficient than conventional pencil and paper tests or non-interactive computer tests. The approach is thus a fast and efficient way of getting a measure of each learner's ability.

In the case of norm-referenced tests, such as the TOEFL test, IRT analysis is used to help establish the equivalency of items used on different forms of the TOEFL test given in various locations towards the goal of reporting a score for the test-taker that is independent of the time, place, or items on the test. The test makers create new items with this objective in mind. Item characteristics are only as important as their contribution to the overall test-taker's score.

In our case however, we employ IRT in a very different way. While the on-line CAT performs as described above to obtain an ability score for the test-taker, our subsequent analysis is much more focused on the item. The response data from the V-Check vocabulary test are used to calibrate the difficulty of each vocabulary word (each item) in our database. New items are drawn from the lexicon, not created by the test makers. From the 20,000 high frequency words we test, any one respondent will typically see no more than 30 actual items during their testing session. However, over time, IRT allows us to establish a precise measure for the difficulty of each item in our item pool for a given population group. The comprehensive testing of a large number of words allows us to create a rank order of the vocabulary items by their difficulty (Culligan, 2008). The CAT gives us the ability to both ascertain each respondent's lexical ability,

as well as to statistically predict which specific words are likely known, and not known for each ability score. In other words, V-Check measures each test-taker's vocabulary ability and then, with a high degree of probabilistic accuracy, identifies the words they know, and more importantly, allows us to identify the high frequency words that they don't know. One of the most unique and useful aspects of the V-Check test is that since the IRT analysis allows us to predict which specific words are already known by the learner, each student who takes the test is able to receive their own personalized list of next most important high frequency words for study.

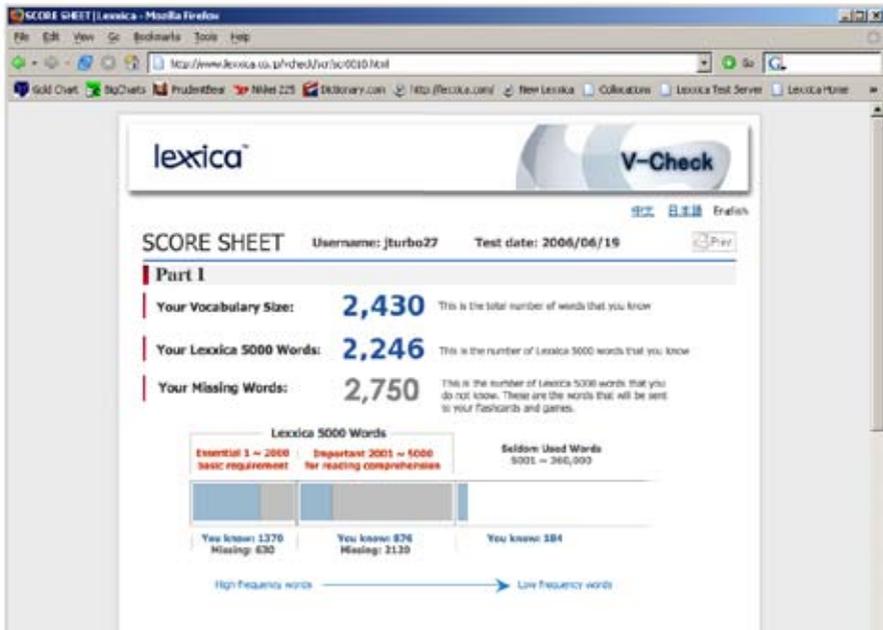


Figure 1: Sample score sheet

In the V-Check test, the computer presents vocabulary words and the respondent is asked, "Do you know this word?" The respondent gives either a "Yes" or "No" response. This type of test is referred to as a Lexical Decision Task or a Yes/No test. Research has shown that the Lexical Decision Task approach is one of the most highly reliable and statistically valid forms of vocabulary testing. (Culligan, 2008; Harrington, 2006; Meara, 1992; Meara & Buxton, 1987). One important benefit of the Lexical Decision Task is that learners are able to respond to more items in a given amount of time than in traditional types of vocabulary tests. The V-Check test finishes in six to 15 minutes, dependent on the response pattern of the test taker. Based on the precepts of Signal Detection Theory, a number of items presented to the test-taker are pseudowords, which are also known as nonsense words or non-words. Pseudowords are strings of letters that have the characteristics of English words but do not have any meaning. These pseudowords serve two purposes. The first is used to control for random behavior on

the part of the test-taker. If a test-taker responds "Yes" to a high proportion of pseudowords (known in the literature of SDT as the "False Alarm" rate), the test is usually rejected. The second purpose is to adjust the score commensurate with the False Alarm rate. In V-Check, the False Alarm rate is used in three ways. If the test-taker responds "Yes" to a predetermined number of pseudowords, the test will automatically terminate. Second, the amount of information accumulated by the test-taker is decreased with each "Yes" response, thus increasing the number of items necessary for the successful completion of the test. Third, the False Alarm rate affects the item selection, reducing the difficulty of the next item. In this way, it works similarly to a "No" response.

As can be seen from figure 1, results with Japanese test-takers have revealed an interesting aspect of EFL word knowledge among Japanese students – there tends to be significant discrepancies between the order of frequency and of difficulty (Culligan, 2008). This particular student score sheet (Figure 1.) is from a first year university student in Japan. It indicates that he knows more than 2400 words, a score not dissimilar to the research results of Shillaw (1995), and Barrow et al. (1999). This sample V-Check score sheet also shows a very large gap of 630 missing (unknown) words from among the first 2000 most frequent words of English. While this student recognizes 2430 total words of English, he is missing many of the most frequently used words, which greatly limits his reading comprehension ability. This profile is typical of the thousands of Japanese students who have taken the V-Check test. Research has shown that knowledge of the first 2000 most frequent words is crucial to gaining basic proficiency in English. The first 2000 words provide up to 85% coverage of written texts (Nation, 1990, 2001). For most Japanese students, learning their missing high frequency words will be of critical assistance in helping to achieve independence as learners. Interestingly, the score sheet also points to the fact that the student knows quite a few lower frequency words (more than 870 words in the 2001-5000 frequency band) and another 180 words that are beyond even the 5000-word frequency level.

Why do such vocabulary knowledge gaps occur? Although it is not within the specific scope of this article, researchers have indicated two contributing factors, the first being the extreme difficulty of reading texts used in high schools and on college entrance exams, and the second, the undue emphasis that Japan's secondary education system's places on teaching English for the purpose of passing college entrance exams rather than for communication (Browne, 1996, 1998, 2002; Butler and Iino, 1995; Kikuchi, 2006; Kitao and Kitao, 1995).

Direct Study of Vocabulary Through Flashcards and Games

Once a person's missing, unknown, or unclear high frequency words have been identified, how are they learned? Research going back more than a hundred years (Ebbinghaus, 1885; Leitner, 1972; Mondria, 1994; Pimsleur, 1967), has shown that learning new words via spaced repetition of flashcards is one of the most efficient ways to quickly increase one's vocabulary size and to move knowledge of these words from short term to long term memory. In their "hand computer" studies, both Leitner (1972) and Mondria (1994) devised elaborate spaced repetition systems for quickly learning new words. Use of personal computers was not yet widespread at the time they published their studies, so they recommended the use of packs of vocabulary cards and a shoebox divided into multiple slots, with each slot representing a different time interval for review. Although the results of these studies were very promising, keeping track of

the correct time intervals for the review of hundreds of physical cards proved to be too cumbersome and demanding for most learners. Another obvious problem is identifying the words that are necessary to study. The physical process of reading thousands of cards to eliminate the known words and concentrate on the unknown words is time consuming and tedious.

By utilizing electronic applications, many of the weaknesses inherent in working with paper cards and shoeboxes can be eliminated. After determining the lexical ability of the test-taker, the V-Check application communicates the test-taker's ability to the basic learning application, the 'Word Engine.' This Word Engine application eliminates all words that have a high probability of being known, and sends a stream of the most frequent unknown words to other learning applications. We have developed spaced repetition applications, which function autonomously for each individual student. These applications utilize multiple learning games and electronic flashcards, and keep track of every response and interaction regardless of the type of electronic interface (PC or mobile) that the users prefer. The Word Engine automatically prioritizes each learner's flashcards and then, through the use of time tags, recycles the words via a spaced repetition process similar to that outlined by Mondria (1994) to deliver learners the words they need to review at the right time intervals.



Figure 2: Electronic flashcard

Figure 2 shows the front and reverse sides of an electronic flashcard. The front side of the card displays the target word “substitute” along with the question “Do you know this word?” The learner tries to recall the definition, and then checks if they were right or not by clicking

the “Check It” button. If they were correct, and they reply, “Yes”, the word is automatically retired until the spaced repetition timer calls for it to once again be displayed. If the user did not properly know it, and they replied “No”, the word will return to the initial interval slot, and wait its turn to start the process again.

Information appearing on the reverse side of each flashcard is adjustable by the user to include the following options.

- Definitions in English (including different “senses” whenever a word has multiple meanings)
- Definitions in the learner’s first language
- Part of speech
- Sound files with native speaker pronunciations of the words
- Frequent collocations for each target word (based on corpus analysis)
- Sample sentences

At the time of this article’s publication, the Word Engine was capable of testing over 20,000 words. For teaching purposes, our database can supply learners with the first 5000 most frequent words of general English as well as 3000 additional special purpose words that are specific to academic testing purposes. For example, the Word Engine contains words that are infrequent in general English but are more frequent on TOEFL, TOEIC, and private university entrance exams.

Another possible problem with vocabulary flashcards is how to sustain learner motivation. Although research suggests that flashcards are a very efficient way to learn new words (Leitner, 1972; Mondria, 1994; Nation, 1990, 2001), students may lose interest if flashcards are the sole method of doing vocabulary review. There is a rich tradition in the ESL/EFL classroom of using games with a communicative purpose to increase and maintain learner motivation (Ersoz, 2000; Uberman, 1998; Wright, Betteridge, & Buckby, 1984) as well as lower the learner’s affective filter (Asher, 1965, 1977; Dulay, Krashen, & Burt, 1982; Krashen 1985). This body of research informed the development of several interactive vocabulary-learning games induce automaticity (discussed below). The games are integrated with the spaced repetition system such that whether a learner reviews new target words with the flashcards or the games, whenever they correctly recognize the meaning of a word, it will be automatically forwarded to the next stage of the spaced repetition process.

Vocabulary Size and Automaticity

Although there is a growing body of research which has established a relationship between vocabulary size and reading ability (Anderson & Freebody, 1981, Beck et al., 1982, Davis, 1968, Laufer, 1992; Nation, 1990), work done by Daneman and Green (1986) suggests that simply increasing a learner’s vocabulary size will not be sufficient since reading comprehension depends not only on the number of words a learner knows, but also on the speed with which they are able to recall each of the word meanings they have stored in their memory. Eskey (1988) argues that the rapid and accurate decoding of language is extremely important to any kind of reading and especially important for second-language readers. Earlier work by LaBerge and Samuels (1974) also points out that fluent readers tend to be able to automatically recognize most of the words they read. In a discussion of the various processes involved in reading comprehension, Abdullah (1993) argues that it appears humans have a finite amount of processing



Figure 3: Sample game activities

ability and that the automaticity of lexical access can free up cognitive processing capacity which can be devoted to the comprehension of text. In other words, fast decoders of a language will have a better chance to be a good reader. Warrington (2006), describes several activities and strategies to help build the automatic process of word recognition for those who are less proficient at reading including reading aloud, extensive reading, word and definition matching, and reading strategy development.

The two games shown in Figure 3, Sight-Words and Sound-Bubbles, are informed by the research on developing automaticity of word knowledge. In Sight Words, the Word Engine delivers both new unknown high frequency words and words for review after the specified time interval. The student is asked to quickly match the word to the correct response, which may appear in the student's native language or in the target language depending on user's settings. Points are awarded to the student based on the speed of identification, thus encouraging rapid responses and automaticity. In Sound-Bubbles, the student first clicks on a bubble to hear a word pronounced. The student then matches a correct response to each sound bubble. Again, speed is rewarded to encourage rapid recall.

Indirect Development of Vocabulary Size Through Reading and Listening

Extensive reading of graded reading materials has been widely used as a way to increase vocabulary size and to improve both motivation and overall ability in English (Day & Bamford, 1998; Susser & Rob, 1990). With evidence that EFL reading materials are too advanced for learners in Japan, (Browne, 1996, 1998, 2002; Butler & Iino, 1995; Kikuchi, 2006; Kitao & Kitao, 1995), use of graded reading materials has been strongly promoted in recent years in Japan through organizations such as SSS (<http://www.seg.co.jp/sss/>). Until now, most attempts to conduct extensive reading programs have made use of physical books, usually through the creation of graded libraries from which students can borrow books to read at home.

The research on the benefits of reading for the development and deepening of the learner's lexicon motivated the establishment of a system for generating graded reading materials that can be accessed online. With the growth of the e-book market in English speaking countries and more recently in Japan, and a virtual 100% cell phone ownership rate by Japanese college students (Browne et al, 2008) mobile devices and personal computers were the obvious platforms of choice. Rather than trying to duplicate the work done by the major publishing houses in producing book-length graded readers, we have opted instead for the creation of materials on current topics of much shorter length (approximately 1000-1200 words) deemed more appropriate for reading on a cell phone. On-going research will determine the suitability of other genres, as well as how the V-Check ability score can inform the selection of text for the student.

V-Admin: Keeping Track of It All

Although all of the software has been designed for individual learners to self-access in an easy and intuitive way, from the very beginning we realized that cell phone software for measuring and tracking vocabulary development might also be of potential interest to teachers and CALL administrators. To this end, we created the V-Admin administration application, a teacher-centered course management program.

As can be seen in Screenshot 4, the current version of the V-Admin program allows a teacher to track student scores on the V-Check vocabulary test. In order for the program to work, teachers must first log in and create their account. They can then create as many classes as they want. Each time they create a class, a code is generated which the teacher gives to their students. When students log in, they are asked if they have a class code. If they enter the code, the student's V-Check scores will be automatically reported to the teacher. The V-Admin is now in the process of being upgraded so that it can track student progress on all other programs such as the amount of time spent on flashcards, games, graded reading materials, as well as the

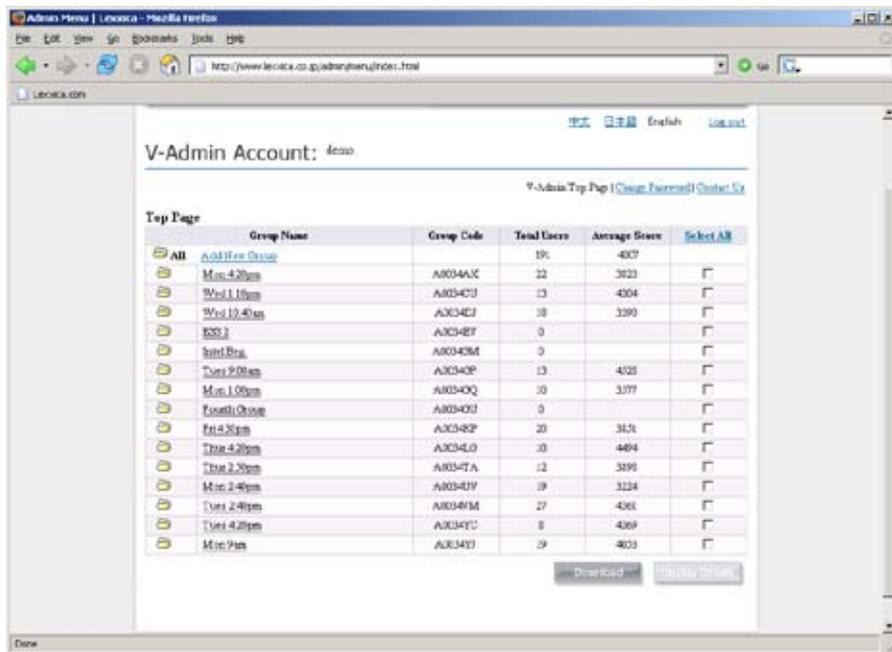


Figure 4:V-Admin screen

number of word encountered. The V-Admin can also generate individualized quizzes based on the words the students are studying for in class testing, thus providing an external check to the activity reports.

After several years of research, software development, and extensive testing, this free suite of programs has only recently become available for online use by students and teachers. With many schools not able to afford adequate CALL facilities to accommodate all their students, or the necessary funds to support teacher training in use of the equipment they do have, we hope this program, combined with the ubiquity of cell phones and personal computers, will help teachers and schools to be able to utilize this modality as a new type of self-access center.

References

- Abdullah, K. I. (1993). Teaching reading vocabulary: From theory to practice. *TESOL Forum*, 31, 10.
- Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. Guthrie (ed.), *Comprehension and teaching: Research reviews* (pp. 77–117). Newark, DE: International Reading Association.
- Asher, J. (1965). The strategy of total physical response: An application to learning Russian. *International Review of Applied Linguistics*, 3, 291–300.

- Asher, J. (1977). Children learning another language: A developmental hypothesis. *Child Development*, 48, 1040–1048.
- Barrow, J., Nakanishi, Y., & Ishino, H. (1999). Assessing Japanese college students' vocabulary knowledge with a self-checking familiarity survey. *System*, 27, 223–247.
- Brown, J. D. (1995). *The elements of language curriculum: A systematic approach to program development*. Boston: Heinle & Heinle Publishers.
- Beck, I.L., Perfetti, C. A., & McKeown, M. G. (1982). The effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, 74, 506–521.
- Browne, C. (1996). Japanese EFL textbooks: How readable are they? *Temple University Japan Working Papers in Applied Linguistics*, 8, 28–41.
- Browne, C. (1998). Japanese high school reading textbooks: Help or hindrance? *Temple University Japan Working Papers in Applied Linguistics*, 12, 1–13.
- Browne, C. (2002). To push or not to push: A vocabulary research question. *Aoyama Ronshu* (11/02), Aoyama Gakuin University Press.
- Browne, C. (2004a). The development of CALL and current options. In S. Fotos & C. Browne (Eds.), *New perspectives on CALL for second and foreign language classrooms* (pp. 3–14). Mahwah, NJ: Lawrence Erlbaum Associates.
- Browne, C. (2004b). *CALL Lab Design 101: Proceed at your own risk* (pp. 1–28). Tokyo: Media no Ibunkakan Eikyoryoku, Aoyama Gakuin University Sokken Research Institute.
- Browne, C., Dias, J., Menish, M., Pagel, J., Shew, P., and Someya, Y. (forthcoming). Faculty and Student-created Audio and Video Podcasts in an EFL Setting. *Aoyama Information Science Journal*, Aoyama Gakuin University.
- Butler, Y.G., & Iino, M. (1995). Current Japanese reforms in English language education: The 2003 Action Plan. *Language Policy*, 4, 25–45.
- Cancino, H., Rosansky, E. J., & Schumann, J. H. (1978). The acquisition of English negatives and interrogatives by native Spanish speakers. In E. Hatch (Ed.), *Second language acquisition* (pp. 207–230). Rowley, MA: Newbury House.
- Carroll, J. B. (1971). Statistical analysis of the corpus. In J. B. Carroll, P. Davies, & B. Richman (Eds.), *The American heritage word frequency book* (pp. xxi–xl). Boston: Houghton Mifflin.
- Chapelle C. (2003). *English language learning and technology*. Philadelphia: John Benjamins Publishing Company.
- Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics*, 5, 161–170.
- Culligan, B. (2008). *Estimating word difficulty using Yes/No tests in an IRT framework and its application for pedagogic objectives*. Unpublished dissertation. Tokyo: Temple University Japan.
- D'Anna, C. A., Zechmeister, E. B., & Hall, J. W. (1991). Toward a meaningful definition of vocabulary size. *Journal of Reading Behavior*, 23, 109–122.
- Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, 25, 1–18.
- Davis, F.B. (1968). Research in comprehension in reading. *Reading Research Quarterly*, 4, 499–545.
- Day, R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge: Cambridge University Press.

- Ducate, L., & Arnold, A. (Eds.) (2006). Calling on CALL: From theory and research to new directions in foreign language teaching. *CALICO Monograph Series, Volume 5*. Texas: Texas State University.
- Dulay, H. C., & Burt, M. K. (1973). Should we teach children syntax? *Language Learning*, 23, 245–258.
- Dulay, H., Krashen, S. & Burt, M. (1982). *Language two*. Oxford: Oxford University Press.
- Ebbinghaus, H. (1964, originally published, 1885). *Memory: A contribution to experimental psychology*. New York: Dover.
- Egbert J. (2005). *CALL essentials: Principles and practice in CALL classrooms*. Alexandria, Virginia: TESOL.
- Ellis, R. (1992). Grammar teaching – practice or consciousness-raising? In R. Ellis (Ed.) *Second language acquisition and language pedagogy*. Clevedon, Avon: Multilingual Matters.
- Ersoz, A. (2000). Six games for the EFL/ESL classroom. *The Internet TESL Journal*, Vol. VI, No. 6. Retrieved October 30, 2008, from <http://iteslj.org/Lessons/Ersoz-Games.html>
- Eskey, D. 1988. Holding in the bottom: An interactive approach to the language problems of second language readers. In P. Carrell, J. Devine, & D. Eskey (Eds.), *Interactive approaches to second language reading* (pp. 93–100). Cambridge: Cambridge University Press.
- Fotos, S. & Browne, C. (Eds.). (2004). *New perspectives on CALL for second and foreign language classrooms*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Goulden, R. Nation, I. S. P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, 11, 341–363.
- Hanson-Smith, E., & Rilling, S. (2006). *Learning languages through technology*. Alexandria, VA: TESOL Publications.
- Harrington, M. (2006). The lexical decision task as a measure of L2 lexical proficiency, *EUROSLA Yearbook*, 6 (1), 147–168.
- Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8, 689–696.
- Kikuchi, K. (2006). Revisiting English entrance examinations at Japanese universities after a decade. *JALT Journal*, 28 (1), 77–96.
- Kitao, K., & Kitao, K. S. (1995). *English teaching: Theory, research, practice*. Tokyo: Eichosha.
- Krashen, S. (1985). *The input hypothesis*. London: Longman.
- LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293–323.
- Larsen-Freeman, D. (1991). Second language acquisition research: Staking out the territory. *TESOL Quarterly*, 25, 315–350.
- Laufer, B. (1992). How much lexis is necessary for reading comprehension? In H. Béjoint & P. Arnaud (Eds.), *Vocabulary and applied linguistics* (pp. 126–132). London: Macmillan.
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English*. Harlow, England: Longman.
- Leitner, S. (1972). *Solernntmanlernen*. Freiburg, Wien, Basel: Herder.
- Levy, M., & Stockwell, G. (2006). *CALL dimensions: Options and issues in computer assisted language learning*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Loschky, L. (1994). Comprehensible input and second language acquisition: what is the relationship? *Studies in Second Language Acquisition*, 16, 303–323.

- Meara, P. (1992). *EFL vocabulary tests*. Swansea: Centre for Applied Language Studies, University of Wales.
- Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests, *Language Testing*, 4, 142–154.
- Mondria, J. (1994). Efficiently memorizing words with the help of word cards and “Hand Computer”: Theory and applications. *System*, 22 (1), 47–57.
- Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, 19, 304–330.
- Nation, P. (1990). *Teaching and learning vocabulary*. New York: Heinle and Heinle.
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics*, 10, 52–79.
- Pimsleur, P. A. (1967). Memory schedule. *Modern Language Journal*, 51 (2), 73–75.
- Sharwood Smith, M. (1981). Consciousness raising and the second language learner. *Applied Linguistics*, 2, 159–168.
- Shillaw, J. (1995). Using a word list as a focus for vocabulary learning. *The Language Teacher*, 19 (2), 58–59.
- Shook, D.J. (1994). FL/L2 reading, grammatical information, and the input-to-intake phenomenon. *Applied Language Learning*, 5, 57–93.
- Susser, B., & Robb, T. (1990). EFL extensive reading instruction: Research and procedure. *JALT Journal*, 2. Retrieved October 30, 2008 from <http://www.kyoto-su.ac.jp/~trobb/sussrobb.html>
- Hirsh, D., & Nation, I. S. P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8, 689–696.
- Trinder, R. (2006). *Language learning with computers: The students' perspective*. Frankfurt: Peter Lang.
- Uberman, A. (1998). The use of games for vocabulary presentation and revision. *Forum*, 36 (1), 20.
- VanPatten, B., & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition*, 15, 225–243.
- Warrington, S.D. (2006). Building automaticity of word recognition for less proficient readers *The Reading Matrix*, 6 (1). Retrieved October 30, 2008 from <http://www.readingmatrix.com/articles/warrington/article.pdf>
- White, L., Spada, N., Lightbown, P., & Ranta, L. (1991). Input enhancement and L2 question formation. *Applied Linguistics*, 12, 416–432.
- Wright, A., Betteridge, D., & Buckby, M. (1984) *Games for language learning*. Cambridge: Cambridge University Press.