

# A Study of the Effects of Digital Interactivity on L2 Vocabulary Acquisition

Thomas MACH

## Abstract

*After presenting an overview of problems involved in the shifting usage of the popular term interactive, this paper reports results from an experiment designed to investigate whether or not digitizing second language vocabulary learning exercises leads to improvements in acquisition. Specifically, the study looked at whether the type of interactivity inherent in digital exercises created with Hot Potatoes software yielded better results than similar exercises provided in print format. Results indicate that the digital exercises did not facilitate greater acquisition than their print-based counterparts. Possible implications of this finding, as well as a discussion of some of the intriguing issues that emerged from the study, are offered.*

In the field of English language teaching (ELT), a select number of terms tend to acquire such positive connotations that they exude an aura of unassailable virtue. Currently, *communicative* is arguably one such term. Perhaps *learner-centered* is another. The term that this study addresses is *interactive*. The fact that certain words seem to have achieved special status can be viewed as a mostly positive development in that it suggests that there is broad agreement between researchers and practitioners regarding what generally constitutes good teaching practice. The danger of such terms, however, is that their popularity causes them to be used in so many contexts for so many different purposes that reaching a consensus about what they precisely mean becomes increasingly difficult. They appear to mean rather different things to different people in the field, but nearly every educator seems to agree that they are positive. Thus, they deserve to be approached with a sufficient level of caution when encountered. If we look closely at how they are used, we may at times find that they are providing attractive cover for otherwise questionable practices or unsubstantiated claims.

## Interactivity

The term *interactive* gained popularity during the rapid rise of the communicative approach in language teaching. In ELT literature from the 1980s, it is not uncommon to

find positive mention of interactive techniques, interactive methods, and interactive lessons. In most of these instances, the interaction being referred to is human interaction. Even when the term was applied to materials, it was usually done so to describe materials that encouraged a greater degree of human interaction.

Now, when we jump ahead to the current decade, we find that the notion of interaction is still prominent in education and still has very positive connotations. However, it is often used to refer to something quite different than before. One typical example, from a book entitled *Managing Technological Change: Strategies for College and University Leaders* (Bates, 2000), suggests that computer-assisted instruction “can be more effective than traditional classroom methods because students can learn more easily and more quickly through illustration, animation, different structuring of materials, and increased control of and *interaction* [italics added] with learning materials” (p. 28).

This example illustrates a shift in focus from human interaction to interacting with inanimate materials. The concept of interactivity is broad, so of course both uses are perfectly acceptable. A problem occurs, however, if readers unquestioningly ascribe to its new usage the deservedly positive connotations it earned when it was used to describe the person-to-person exchanges found in communicative classrooms. It may cause us to simply accept claims that interactive materials are superior to other materials without evidence, when instead we ought to be rigorously testing such assumptions.

This study was designed to test the assumption that interactive materials delivered in a digital format are more effective than their traditional, non-interactive cousins. Specifically, it looks at the effects of interactivity on students studying second language (L2) vocabulary. Since the early 1990s, the ELT field has trended towards greater acknowledgment of the role of vocabulary in L2 development. This increased attention to vocabulary has even produced its own word-centered theoretical approach to language pedagogy, the lexical approach (Lewis, 1993). While the trend has contributed to a rise in the number of major studies that try to identify effective L2 vocabulary instruction techniques, the amount of studies that attempt to examine the efficacy of computer-based materials for vocabulary learning is still relatively small (see Nikolova, 2002, for a list of those that do). Nevertheless, with the increasing availability of easy-to-use software for creating digital language exercises and the growing number of computer classrooms in schools, digital materials are no doubt being used more and more in language teaching. As the previous quote from Bates (2000) illustrated, digital learning materials are regularly touted as being more interactive than their predecessors, and thus they are generally assumed to be an improvement over traditional print materials. The purpose of this study, then, was to look at whether digital vocabulary learning materials do indeed yield greater results than typical print ones. The primary research question can be stated as follows: Does the use of digital, web-based, “inter-

active” vocabulary materials promote greater acquisition gains than similar materials provided in traditional print format?

## **L2 Vocabulary Instruction**

How does lexical growth typically occur? Evidence suggests that much of it is the result of incidental vocabulary learning that occurs during the reading process itself. Studies conducted in both L1 situations (e.g., Nagy, Herman & Anderson, 1985) and L2 environments (e.g., Day, Omura & Hiramatsu, 1991) point to the generally positive lexical effects of reading. Clearly, the more reading that learners are engaged in, the more they are exposed to words in natural and meaningful contexts. This exposure is beneficial for deepening lexical knowledge, although there is certainly a degree of individual variation based on reading skills and attitudes toward reading.

It is likely that the appearance of studies showing a positive role for reading in vocabulary development has led some L2 practitioners to conclude that reading alone is enough for sufficient lexical growth and that direct instruction of vocabulary is relatively unproductive. More recent trends in vocabulary acquisition research, however, generally suggest that it would be a mistake to completely discard direct instruction (Boyd Zimmerman, 1997; Paribakht & Wesche, 1997). Perhaps one of the most pressing questions now is not whether or not direct instruction is justified, but rather what kind of direct instruction provides the best results when coupled with a strong reading component.

Boyd Zimmerman’s (1997) study found that students in a reading class consisting of a combination of interactive vocabulary instruction and self-selected reading made greater vocabulary acquisition gains than those in a class that combined reading skills instruction with self-selected reading time. Her study, though conducted in the late 1990s, did not include computer-based materials in its operationalization of the concept of interactivity. Instead, the author adapted Nagy and Herman’s (1987) list of five parameters for interactive vocabulary instruction:

1. multiple exposures to words;
2. exposures to words in meaningful contexts;
3. rich and varied information about each word;
4. establishment of ties between instructed words, student experience, and prior knowledge; and
5. active participation by students in the learning process. (Boyd Zimmerman, 1997, p. 125)

In the context of our increasingly digitized educational environment, it is useful to consider which of the listed items might be better facilitated by computer-mediated instruction. Regarding the first one, multiple exposures, it seems clear that digital materials have the edge over their print counterparts. It is possible to create materials that jumble items anew each time an exercise is accessed, and that delete previously entered answers whenever a refresh button is clicked. This is a much more efficient way of providing multiple exposures to target words than copy machines and pencil erasers are able to offer.

It is more difficult, however, to draw clear conclusions about the second, third, and fourth parameters. As for the second one, though a digital format makes it easier to provide rapid access to a diversity of contexts, there is nothing about the medium itself that makes those exposures inherently more meaningful than the ones that can occur in non-digital formats. Regarding the third item on the list, the information given about words in a digital format may potentially be more varied than that which print materials can provide, but it is difficult to claim that it is necessarily richer as well. In a regular classroom, vocabulary instruction is likely to involve framing provided by the teacher to a greater degree than in computer-centered classrooms, and the teacher's familiarity with students may allow her to make better judgments about how much information is appropriate and what kind of information is comprehensible. This issue of teacher involvement also affects the fourth parameter: A classroom teacher would likely have at least as good a sense as a computer might about the typical experiences and prior knowledge of a particular group of students, and she could presumably use what she knows in order to help students make connections between experiences, prior knowledge, and instructed words.

Finally, any consideration of the final parameter depends greatly on what is meant by "active participation." If it includes the type of control and manipulation of materials that a keyboard and mouse enable, then students working with interactive digital materials are indeed active participants in the learning process. On the other hand, if it refers to the sort of negotiation of meaning with others in the room and collaborative efforts to find answers that typically occur in communicative classrooms, it would seem that non-digital learning environments might have the upper hand. In sum, while it seems reasonable to argue that a digital format provides the best conditions for the first and most basic of the parameters to be fulfilled, there are serious doubts about whether it truly offers the best environment for the remaining four aspects of interactive vocabulary learning as originally defined by Nagy and Herman.

Given the uncertainty about whether the types of interactivity provided by digital materials are similar enough to our traditional notions of interactivity to warrant assumptions of increased pedagogical effectiveness, this present study attempted to iso-

late the effects that a common type of interactive digital materials had on a particular group of English learners.

## Method

### *Participants*

The participants for this study were university students from various departments enrolled in two sections of an elective intermediate-level English reading course. Their only requirement for entering the course was having received passing grades in their three elementary-level English courses. Classes for one of the sections were held in a computer-assisted language learning (CALL) classroom, and students in this class became the experimental group. The other section met in a regular classroom without computers, and this group of students were treated as the control group. From each group, a number of participants were eliminated from the study due to excessive absences, missed tests, or dropping out of the course. From an original total of 32 students enrolled in the experimental group, 26 have been included in the study. The control group had a higher rate of attrition: 17 students have been included from an original total of 30 enrollees.

### *Digital Materials*

The web-based materials used in this experiment were created with the well-known *Hot Potatoes* program. *Hot Potatoes*, marketed by Half-Baked Software, is a package of applications developed by researchers at the University of Victoria that anyone can download in order to create a variety of language learning exercises. It is free for educators, and thus widely known and appreciated among CALL practitioners. Also, teachers who do not consider themselves computer experts find it attractive because it is relatively easy to use. The *Hot Potatoes* website explains that exercises created with the software provide interactivity: “The *Hot Potatoes* suite includes six applications, enabling you to create *interactive* [italics added] multiple-choice, short-answer, jumbled-sentence, crossword, matching/ordering and gap-fill exercises for the World Wide Web” (2005, “What is *Hot Potatoes*?” section, para. 1).

At language teaching conferences, as well as in journals and newsletters aimed at teachers, it is not uncommon to come across enamored users of this software who eagerly spread the word about its usefulness and effectiveness. Here is one such example from a glowing review that appeared in *Language Learning & Technology*:

The *Hot Potatoes* program, which consists of modules for creating six different types of exercises, is an excellent resource for creating on-line, *interactive*

[italics added] language learning exercises that can be used in or out of the classroom. These types of exercises can be especially useful in language learning laboratories with Internet access, or for remote learning. When matched with both appropriate content and motivated students, *Hot Potatoes* exercises seem likely to promote second language acquisition. (Winke & Macgregor, 2001, p. 32)

Beyond its appearance in this excerpt, the term *interactive* appears six more times in the short article, indicating that the reviewers clearly see interactivity as one of the most salient and attractive features of the software. Given the general consensus in CALL circles that *Hot Potatoes* exercises are indeed interactive, they were deemed appropriate for this experiment and, just as importantly, a worthwhile study tool for the students enrolled in the course.

#### *Target Vocabulary*

The vocabulary words taught and tested during the course were taken from the Academic Word List (AWL) developed by Coxhead (2000). The AWL has been derived from a 3.5 million-word corpus of academic texts and consists of the most frequent 570 word families excluding the 2,000 most frequently occurring English words as found in West's (1953) General Service List.<sup>1)</sup> For the purposes of the AWL, a word family is defined as a stand-alone stem (e.g., *structure*) along with its most frequently occurring inflections and other affixes (e.g., *structures*, *structural*, *structured*, *unstructured*, *restructure*, *restructuring*). Though the concept of a word family is presented in linguistic terms, there is psycholinguistic evidence suggesting that the mental lexicon groups known members of word families together, and this serves as a psychological rationale for treating the word family as a unit when teaching vocabulary (Schmitt & Boyd Zimmerman, 2002).

Academic vocabulary was deemed an appropriate learning goal for this course because the students were likely to already have passable knowledge of the 2,000 most frequent English words, but unlikely to have previously studied many of the sub-technical words which make up the AWL. The term *sub-technical* differentiates general academic vocabulary from the technical terms applicable only to specialized fields. It also suggests a somewhat more formal register than is typical of general word lists. Nation (2001) cites a number of studies that have found that English learners tend to be less familiar with sub-technical terms than with the technical terms associated with their fields. It seems that sub-technical academic vocabulary often slips through the cracks of language courses, perhaps because it tends to be less salient than technical terms and more abstract than general vocabulary.

Table 1  
*Coverage Data for Three Types of Texts*

Levels	Academic Text	Newspapers	Conversation
1 <sup>st</sup> 1,000	73.5%	75.6%	84.3%
2 <sup>nd</sup> 1,000	4.6%	4.7%	6 %
Academic	8.5%	3.9%	1.9%
Other	13.3%	15.7%	7.8%

Table 1, adapted from Nation (2001), puts academic vocabulary in perspective regarding the amount of coverage it provides in particular genres.<sup>2)</sup> While academic words comprise just below 2% of conversations, their presence doubles in newspapers and quadruples to more than 8% in academic texts. Though the AWL would probably be an inappropriate source of vocabulary for most conversation courses, Table 1 suggests that it can be very useful for typical university-level reading courses. Without this data, it might be logical to assume that the best place to turn after the second 1,000 most frequent words are mastered is to the third set of 1,000 most frequent words. But in the academic genre, the third 1,000 words provide only 4.3% coverage (Nation, 2001) compared to the 8.5% coverage offered by an academic list. In fact, combining and averaging the data for academic texts and newspapers shows that academic words are more common than even the second most frequent 1,000 English words (6.20% to 4.65%), suggesting that students ought to study academic vocabulary soon after mastering the 1,000 most common English words if fluent reading of non-fiction is a goal.

The AWL is derived from the words that occur most frequently over a wide range of texts covering humanities, science, law, and commerce disciplines, and this range overlaps perfectly with the majors of the students enrolled in the reading courses of this study because it precisely mirrors the types of departments found at the university where the study has been conducted.

The 570 AWL word families are grouped into ten sublists according to frequency. Given the time constraints of the course, this study made use of only the first five sublists covering the 300 most frequently occurring AWL word families.

### *Procedure*

The reading classes met once a week for 90 minutes throughout the academic year, resulting in 26 total classes for both the experimental and control groups. In each class, approximately 30 minutes was devoted to vocabulary study. Nearly all of the time devoted to vocabulary study consisted of individual or paired work on practice

exercises involving the target AWL terms. The major difference was that the experimental group did the exercises entirely online, whereas the control group worked with print versions of the same questions.<sup>3)</sup>

The digital and print versions for most question types (e.g., multiple-choice exercises, cloze passages, crossword puzzles) were nearly identical in what they asked the students to do. The noticeable difference was that the experimental group was clicking or selecting while the control group was circling or writing their answers. The matching exercises differed somewhat more because students in the experimental group were clicking and dragging chunks of text across their screens while the control group was drawing lines or filling in blanks.

The two groups differed greatly, however, in the way that they received corrective feedback. The digital exercises provided immediate feedback to individuals whenever an on-screen *check* or *hint* button was clicked. The students in the control group, in contrast, had to wait for their teacher to provide the correct answers. This resulted in somewhat more individual study time for the experimental group. They were in control of the whole 30 minutes set aside in each class for vocabulary study, whereas about ten of those 30 minutes in each of the control group's classes became a more teacher-centered time for answer confirmation.

The mention of teacher-centered time brings up one more important aspect in which the groups differed from each other. Though negative connotations are usually associated with the term *teacher-centered*, when compared to computer-centered learning, we can say that a teacher-directed activity usually involves more human interaction. Also, when students in both groups worked on exercises in class, they were nearly always given the option of working individually or working with a partner as long as they used English when working collaboratively. The students in the traditional classroom tended to work in pairs more often than the ones in the computer classroom did. Some control-group pairs collaborated quite actively while others preferred to just consult each other occasionally. The students using digital materials, in contrast, generally became solely focused on their individual computer screens. Given the richly interactive nature of their materials, they apparently saw little point in consulting with partners. The digital materials also included links to online dictionaries and a thesaurus, thereby increasing student autonomy. In short, as a result of adapting to their different learning environments, the print group engaged in more human interaction while the digital group interacted more with their materials.

An unannounced pretest was administered at the beginning of the course to both groups in order to check existing knowledge of the target AWL vocabulary subsets. The same test was administered as a posttest, again unannounced, eight months later in the second to last class meeting in order to measure any gains. In addition, two



progress tests were given during the course. They were announced and students were encouraged to study for them and told that scores would affect their course grades. The first progress test covered AWL subsets 1 and 2 and was administered in the 12<sup>th</sup> week of class. The second one covered AWL subsets 3 and 4 and was administered during the 21<sup>st</sup> week of class.

### *Assessment*

For the purposes of this study, a test called the Vocabulary Level Check Test (VLCT) was designed to serve as a comprehensive measure of familiarity with the first five subsets of AWL words. The VLCT measures both receptive and productive knowledge, and gives an indication of knowledge depth for certain items. It consists of 60 total items divided into three sections. Twelve words from each of the five subsets were randomly chosen as items and evenly distributed in each of the three sections.

The first section consists of 20 multiple-choice questions as a measure of receptive, or passive, vocabulary knowledge. While generally recognized as a valid format for assessing receptive vocabulary knowledge, one often heard criticism of multiple-choice items is that they may be testing learners' knowledge of distractors rather than target words (Read, 2000). That is, it is possible to choose the correct answer even if it is unfamiliar as long as all of the distractors can be ruled out. On the VLCT, however, all of the distractors for each item were chosen from the same AWL subset as the target word. Thus, getting an answer correct via ruling out distractors is still a measurement of familiarity with the AWL. In a sense, since each of the multiple-choices questions offer four possible answers, these 20 items can actually be said to measure general familiarity with 80 AWL words.

The second section consists of 30 partial blank-filling items in which most of the target word is removed from a sentence but two or three of the initial letters remain in order to ensure that only the intended AWL word can be elicited as a correct fit. Figure 1 shows some typical examples of this type of item.

1) How many courses did you <u>REG</u> _____ for this year?
2) Please <u>IND</u> _____ your first choice with a circle.
3) I <u>PU</u> _____ my lunch at <i>Family Mart</i> for ¥580 today.

*Figure 1: Examples of partial blank-filling items*

Because learners need to produce the remainder of the target words themselves, this item type goes beyond receptive knowledge. However, it is not a completely open measure of productive ability. The initial letters serve to limit the number of possible correct answers and they may facilitate recall. Nevertheless, the partial blank-filling item type can be said to assess *controlled productive knowledge* (Laufer, 1998), and Laufer and Nation (1999) argue that scores based on such items can be interpreted as measures of the number of words available for productive use.

The final section of the test consists of ten questions and was designed to measure depth of vocabulary knowledge. For this purpose, a Japanese variant of Paribakht and Wesche's (1997) Vocabulary Knowledge Scale (VKS) was developed. The VKS asks learners to locate their knowledge of a target word within the 5-level scale shown in Figure 2.

- 1) I don't remember having seen this word before.
- 2) I have seen this word before, but I don't know what it means.
- 3) I have seen this word before, and I think it means \_\_\_\_\_.
- 4) I know this word. It means \_\_\_\_\_.
- 5) I can use this word in a sentence: \_\_\_\_\_. (Write a sentence.)  
(If you do this section, please also do section 4)

*Figure 2: Paribakht and Wesche's (1997) VKS elicitation scale*

The first two levels of the scale rely on self-reporting, levels three and four require supporting evidence, and level five, the deepest of the levels, requires full productive capability. The Japanese version designed for the current study and shown in Figure 3 follows the same basic format. Note that test-takers are allowed to provide evidence for level three and four responses in either Japanese or English.

<p>1) この単語は見たことがない。</p> <p>2) この単語は見たことはあるが、意味を知らない。</p> <p>3) はっきり分らないが、たぶんこの単語の意味は _____ と思う。 (日本語、英語どちらでも良い)</p> <p>4) この単語を知っている。意味は _____ である。 (日本語、英語どちらでも良い)</p> <p>5) この単語を使って、英文を一文書け。(5に答える者は必ず4にも解答せよ。)</p> <p>_____。</p>
--

Figure 3: A Japanese version of the VKS elicitation scale

The scoring procedures for the first two sections of the VLCT are relatively straightforward, while the scale used for scoring the VKS section is considerably more complex. The item weighting and scoring system for the entire 60-item, 120-point test is presented in the Appendix.

In addition to the VLCT that was used for pretest and posttest purposes, two smaller scale progress tests were also designed for this study. Both of them consisted of four sections: a simple matching section and multiple choice section for assessing receptive knowledge, a partial blank-filling section for measuring controlled productive skills, and a section in which test-takers were confronted with a list of target words in order to freely write sentences to demonstrate their productive mastery of the self-chosen terms.

Finally, two questions that specifically addressed the experience of studying AWL vocabulary were added to the year-end course evaluation form that students answered anonymously.

### Analysis

In order to confirm that the experimental group was comparable to the control group before the treatment, a two-sample t-test was conducted on the VLCT pretest raw scores. It revealed no significant difference between the groups at the  $p < .05$  level in terms of their overall knowledge of the target vocabulary,  $t(41 df) = 0.42$ . T-tests were also separately performed on the three sections of the pretest, all yielding results that were not significant,  $t(41 df) = 0.44, 0.79,$  and  $0.39,$  respectively;  $p < .05$ . Thus, we can be confident that receptive and productive knowledge, as well as depth of knowledge of the target vocabulary were all comparable between the two groups when the course began. Consequently, it is not unreasonable to assume that any gains in vocabulary scores between the pretest and the posttest resulted from the treatment

rather than from preexisting differences.

Paired means t-tests were performed using the pretest and posttest scores of the VLCT to determine whether the digital group and/or the control group made significant improvements in their knowledge of the target vocabulary by the end of the course. Also, t-tests were used to determine whether the two groups differed from each other on the VLCT posttest, as well as on the two progress tests that preceded it.

Finally, to get some idea of what students in the two groups thought about the materials they used, the means and standard deviations for their responses to the questionnaire items that asked whether studying AWL vocabulary was enjoyable and helpful were calculated and are reported in the following section.

## Results

Results of paired means t-tests comparing students' VLCT pretest and posttest scores show that significant gains ( $p < .05$ ) were made by both the experimental group [ $t(25 df) = 10.08$ ] and the control group [ $t(16 df) = 10.43$ ] during the course of the year. As Figure 4 illustrates, both groups made about a 16% jump in their knowledge of the target vocabulary. They both began the course with test scores averaging just short of 35% and both finished the course with averages close to 50%. Also, as Figure 4 graphically suggests and a t-test confirms, the two groups were not significantly different from each other at the end of the year,  $t(41 df) = 0.20$ ;  $p < .05$ . In other words, vocabulary acquisition gains by the digital group were no greater than those of the control group despite their different treatments.

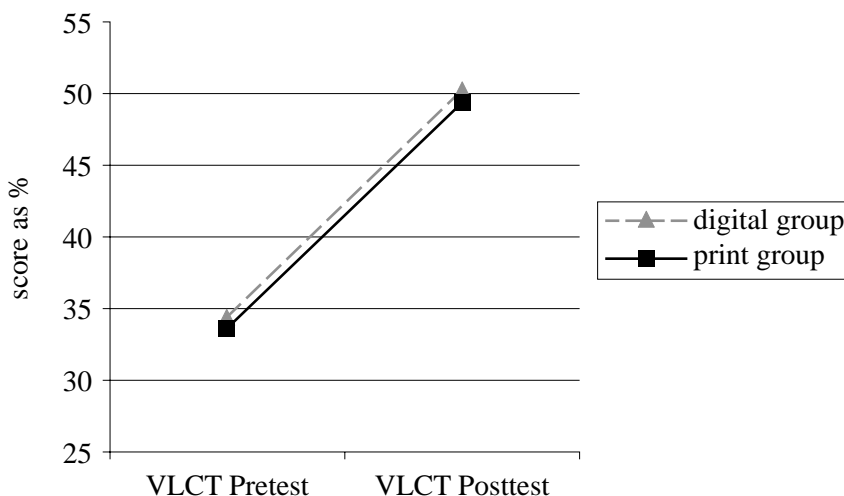


Figure 4: Comparison of VLCT mean scores

The data in Figure 5 presents a somewhat more complex picture. Like Figure 4, Figure 5 shows the mean scores for the VLCT pretests and posttests, but it also shows the means for the two intervening progress tests.

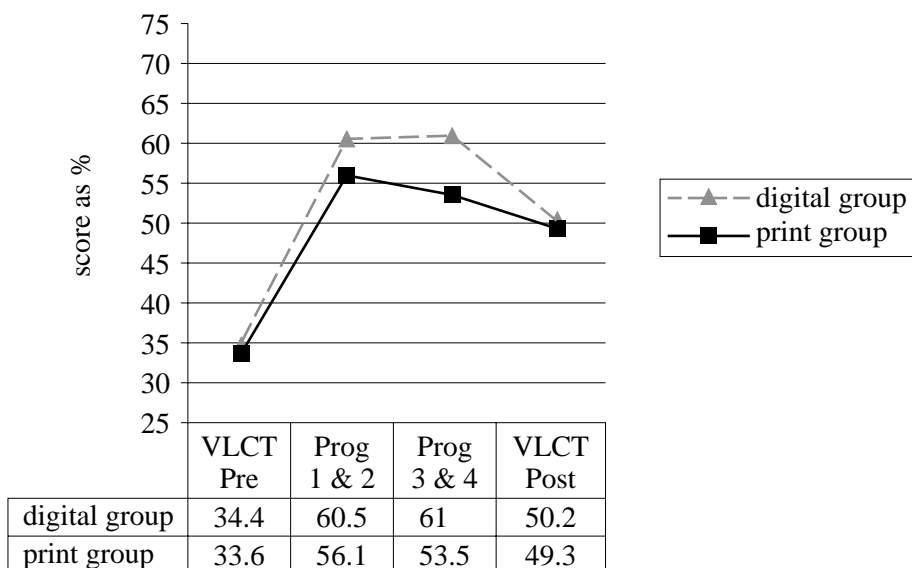


Figure 5: Comparison of mean scores for all tests in the study

Keep in mind that the VLCT and the progress tests differed not so much in item type, but in how they were administered: Students were not told of the VLCT tests before taking them, but they knew the dates of and were given time to prepare for the progress tests. When simply comparing the means, the digital group appears to have outperformed the control group in both of the progress tests. In fact, two-sample t-tests indicate that the difference in means is not significant for either of the progress tests [ $t(41 df) = 0.80$  and  $1.19$ , respectively] at the  $p < .05$  level, so it would be inappropriate to suggest that an actual difference has been observed. Nevertheless, the apparent discrepancy observed in Figure 5 suggests that this small study might be pointing to a potential difference that a more rigorous study with a larger sample size could uncover: namely, the possibility that digital materials enable stronger performance on announced tests that students are able to prepare for, but not on unannounced tests that can be said to more accurately measure acquisition levels since they avoid any effects from short-term cramming.

Table 2 shows mean scores and standard deviations for the questions on the course evaluation questionnaire that addressed AWL vocabulary study. Note that for each question, although both groups are well into the positive side of the scale, the students

who used digital materials responded more positively than those who used print ones: The digital group enjoyed studying the target vocabulary more and had a greater belief in the ability of the materials used in class to improve their English.

Table 2  
*Questionnaire Responses*

Question	Digital Group		Print Group	
	<i>M</i> *	<i>SD</i>	<i>M</i>	<i>SD</i>
I enjoyed studying AWL vocabulary in this class	3.88	.76	3.53	1.01
I believe studying AWL in this class has improved my English ability	4.38	.49	4.06	.74

\*Note: Means calculated from responses given on a scale from 1 (*strongly disagree*) to 5 (*strongly agree*).

Table 2 also shows that, for each question, the digital group's means have smaller standard deviations. This indicates a more narrow range of responses, thereby suggesting a comparatively high degree of consensus among students in the digital group when faced with these questions.

## Discussion

The primary reason for conducting this study was to try to answer the following question: Does the use of interactive, digital vocabulary materials promote greater acquisition gains than similar materials provided in traditional print format? As Figure 4 has shown, the answer appears to be negative. Both groups made remarkably similar gains. Whether the failure of students using digital materials to outperform those using print ones is seen as a disappointment or not depends largely on initial expectations. It is probably an unwelcome result to avid proponents of the type of interactivity that typical digital materials provide. On the other hand, for teachers who fear that more is lost than gained when making the switch to digital, this result could be seen as comforting. That is to say, at least in the case of L2 vocabulary acquisition, students in a digital environment appeared to perform no worse than students learning in the more traditional environment that is so familiar to most of us. However, given the at times hyperbolic claims regarding the effectiveness of the computer as a learning tool and the high costs that are often entailed, it seems safe to say that CALL materials are generally expected to be *more* effective rather than *just as* effective. Nevertheless, although

this study investigated just one little corner of the overall e-learning movement, it can offer no support for the increasingly commonplace educational assumption that digital interactivity provides improved learning results.

The results of this study may, however, point toward an intriguing explanation of why such an assumption of digital superiority exists in the first place. Figure 5 appears to show the digital group scoring higher on announced progress tests than their print-based peers. Typical assessment in most language courses has much more in common with the conditions under which these progress tests were administered than with the unannounced VLCT pretest and posttest. Typically, students are told of an upcoming test, told what might be on the test, and told that it will affect their grades. If digital materials help students perform better in such a situation, it is not hard to imagine how assumptions regarding the superiority of digital materials might take root before being rigorously investigated. Could it be that digital materials lead to greater short-term learning bumps than print ones do, and thereby cause us to wrongly assume that they facilitate greater long-term acquisition? Of course, larger studies would have to be undertaken in order to prove this point. Also, the differences in progress test mean scores in the present study were not statistically significant, so it is important not to overstate the possibility. However, a thorough addressing of this issue may be one of the more interesting avenues to follow for further research.

Another intriguing path of potential inquiry is suggested by the results in Table 2. Despite the fact that the students in the digital group did not outperform their peers on the posttest, they had a greater *belief* in the efficacy of the studying they did when compared to the responses provided by the print group. In addition, as a comparison of the standard deviations shows, the digital group had higher internal agreement when providing their positive feedback. It is also worth noting that providing feedback on a questionnaire is not the only option students have for evaluating a course that they have electively enrolled in. Some of them, if displeased for any reason, simply stop attending class. As mentioned at the outset, the control group in this study had a higher rate of attrition: 17 of 30 students (56.7%) remained at the end of the course as opposed to 26 of 32 students (81.3%) remaining in the digital group. We can only speculate as to what caused this noticeable gap in attrition rates, but given the data in Table 2 which shows that the digital group both enjoyed their AWL study more and had a greater belief in the effectiveness of their materials, it is logical to surmise that the materials used by each group might be one of the factors that accounts for the difference in attrition. This suggests that, even if digital materials do not lead to better learning results, they may still be worthwhile if they cause students to stay in class and take a greater interest in their learning.

Finally, the digital group's failure to acquire more vocabulary than the control

group may not be such a surprise after all if we remain duly skeptical about what passes for “interactive” in the current state of popular digital materials. As mentioned earlier, the way that some CALL advocates tend to use the term *interactive* is quite different from how it was used in the ELT field when it first earned such positive connotations. When we attempt to apply to typical digital materials any thorough definitions of interactivity as it has traditionally been understood, such as Nagy and Herman’s (1987) parameters for interactive vocabulary instruction, it is not at all clear that the digital environment provides much of an improvement. It seems we are in need of a deeper understanding of precisely what unprecedented learning benefits are offered by our new digital medium, and of whether or not those benefits are inherently associated with the new ways in which learners can interact with the materials.

*Hot Potatoes* exercises were used in this experiment because of their popularity and widespread use in ELT. However, it is important to note that their popularity might be due in large part to their inherent familiarity. Regardless of how effective we consider crossword puzzles, matching exercises, and cloze passages to be, they can hardly be considered groundbreaking. They were developed at a time when the print medium had no rivals, and they are familiar to any language teacher. *Hot Potatoes* software allows us to repackage familiar learning activities for a new medium, and the resulting exercises, thanks to the novelty of the medium, tend to come across as fresh and innovative. The way in which learners interact with the exercises is indeed new, but the essential nature of the learning activities themselves has not undergone any significant changes. It may be that the real promise of the digital medium will not be realized until new and effective learning activities that have no direct precedent in the print medium are conceived of, developed, and gain wide acceptance. Popular software programs like *Hot Potatoes* are useful in that they serve as an unimposing bridge from print to digital, but they stop short of offering a fundamental re-conceptualization of what might be pedagogically possible when we learn to fully exploit the digital medium. When the act of manipulating a keyboard and mouse rather than a pencil and eraser in educational settings loses its lingering novelty and becomes mundanely familiar, we may then finally be able to turn our attention to offering digital materials that are truly innovative and superior.

### Notes

- 1) The entire Academic Word List, as well as information about how it was developed, is available at the following URL: <http://www.vuw.ac.nz/lals/research/awl/awlinfo.html>
- 2) The data in the *Academic* row of Table 1 is derived from Xue and Nation’s (1984) University Word List, a predecessor of the AWL. In fact, Coxhead (2000) claims that the AWL provides an even more impressive 10% coverage in academic texts.



- 3) The digital exercises used in this study can be viewed at the following website: <http://www.kilc.konan-u.ac.jp/~mach/awl/>

## References

- Bates, T. (2000). *Managing technological change: strategies for college and university leaders*. San Francisco: Jossey-Bass.
- Boyd Zimmerman, C. (1997). Do reading and interactive vocabulary study make a difference? An empirical study. *TESOL Quarterly*, 31, 1, pp. 121-140.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 2, pp. 213-238.
- Day, R.R., Omura, C., & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language*, 7, 541-549.
- Hot Potatoes. (n.d.). Retrieved November 6, 2005, from <http://web.uvic.ca/hrd/halfbaked/>
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: same or different? *Applied Linguistics*, 19, 255-271.
- Laufer, B. & Nation, I.S.P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16, 33-51.
- Lewis, M. (1993). *The lexical approach: the state of ELT and a way forward*. Hove, England: Language Teaching Publications.
- Nagy, W.E., & Herman, P.A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M.G. McKeown & M.E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 19-35). Hillsdale, NJ: Erlbaum.
- Nagy, W.E., Herman, P.A., & Anderson, R.C. (1985). Learning words from context. *Reading Research Quarterly*, 20(2), 233-253.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nikolova, O. (2002). Effects of students' participation in authoring of multimedia materials on student acquisition of vocabulary. *Language Learning & Technology*, 6, 1, pp. 100-122.
- 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: A rationale for pedagogy* (pp. 174-200). Cambridge: Cambridge University Press.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Schmitt, N. & Boyd Zimmerman, C. (2002). Derivative word forms: What do learners know? *TESOL Quarterly*, 36, 2, pp. 145-171.
- West, M. (1953). *A general service list of English words*. London: Longman, Green.
- Winke, P. & Macgregor, D. (2001). Review of *Hot Potatoes*. *Language Learning & Technology*, 5, 2, pp. 28-33.
- Xue, G. & Nation, I.S.P. (1984). A university word list. *Language Learning and Communication*, 3, 215-229.

## Appendix

### VLCT Scoring Scale

#### Part 1: Multiple choice

- 4 items from each sublist (or 16 if including distractors)
- 20 items  $\times$  1 = 20 total points

#### Part 2: Partial blank-filling

- 6 items from each sublist
- 30 items  $\times$  2 = 60 total points
- Scoring:
  - 2.0 – correct word (simple spelling mistakes not deducted if form, tense and aspect are all decipherable)
  - 1.5 – correct word; faulty tense or aspect
  - 1.0 – correct word family; wrong form
  - 0.5 – correct word family; invented form
  - 0 – wrong word; wrong family

#### Part 3: Japanese version of the Vocabulary Knowledge Scale

- 2 items from each sublist
- 10 items  $\times$  4 = 40 total points
- Scoring:
 

A all	0
B all	1
C incorrect	1
C partially correct	1.5
C correct	2
D incorrect	1
D partially correct	1.5
D correct	3
E incorrect; D incorrect	1
E incorrect; D partially correct	1.5
E incorrect; D correct	3
E partially correct; D incorrect	1
E partially correct; D partially correct	2
E partially correct; D correct	3.5
E correct; D incorrect	1.5
E correct; D partially correct	2.5
E correct; D correct	4