

Does *Word Coach* Coach Words?

TOM COBB

Université du Québec à Montreal, Canada

MARLISE HORST

Concordia University, Canada

ABSTRACT

This study reports on the design and testing of an integrated suite of vocabulary training games for Nintendo™ collectively designated *My Word Coach* (Ubisoft, 2008). The games' design is based on a wide range of learning research, from classic studies on recycling patterns to frequency studies of modern corpora. Its general usage and learning effects were tested over a four-month period, with fifty age and level appropriate Francophone English as a second language learners in a Montreal school. A battery of observational and empirical tests tracked experimental and quasi-control groups' lexical development on the dimensions of form recognition, meaning recognition, free production, and speed of lexical access, as well as features of game use. Two months' gaming coincided with gains in recognition vocabulary normally achieved in one to two years, longer oral productions, reduced code switching, and increased speed of lexical access. Further questions are raised about the prior knowledge *Word Coach* assumes, the importance of post-game follow up, and the future of commercial gaming in language learning.

KEYWORDS

Vocabulary Acquisition, Research-based Instruction, Data-driven Learning, Corpus-based Learning, Computer-assisted Language Learning, Game-based Learning

INTRODUCTION

Video games occupy more and more of the time and attention of school age learners, with an effect on learning that is presently little known. Arguments in principle for the learning power of such games are many (Gee, 2003; Pensky, 2006), but empirical investigations few. This paper describes the design and testing of a video game focused on the specific goal of vocabulary expansion. *My Word Coach* (Ubisoft, 2008, hereafter *Word Coach*) is a vocabulary training system whose goal is to help either first-language (L1) or second-language (L2) learners of English grow their recognition lexicons systematically by meeting words that are new to them in an integrated suite of word games. Each game focuses on one or more aspects of word knowledge, such as form recognition, meaning association, or lexical access, and each word is met in several games. Cobb was also the linguistic consultant for its design and development (although he was not privy to the precise manner in which every idea provided was eventually implemented). *Word Coach* was first presented to the world at a vocabulary symposium in 2007 which gathered together large scale projects in which vocabulary research had been deployed to solve "real learners' vocabulary needs."¹

The Vocabulary Problem

Who are real learners and what are their vocabulary needs? Real learners here refers to those learning English for academic or professional purposes whose careers are likely to be substantially affected by knowledge of English or the lack of it. What are the vocabulary

needs of such learners? The vocabulary research literature, since Meara declared vocabulary a “neglected area” in 1980, has become vast and diverse, yet some common themes have emerged. It is now generally agreed that vocabulary acquisition is not as easy as we used to think (Cobb, 2007; Laufer, 2005), that vocabulary is more important in general language development than we used to think, whether in L1 (Bates & Goodman, 1997) or L2 (Barcroft, 2007), and that many language learners at every level have inadequate vocabulary knowledge for the educational and professional tasks they have set themselves, whether in an L1 (Chall & Jakobs, 2003) or L2 (Laufer, 2000). Laufer’s (2000) synthesis of vocabulary size studies of academic learners in seven non-Anglophone countries found such learners are attempting to follow English-medium university courses with an average vocabulary size of 2,100 word families ($SD=977$); this was in contrast to lexical coverage studies that target the vocabulary needed for basic academic reading at more like 8,000 word families (e.g., Nation, 2006). The vocabulary needs of these learners can be stated simply, to learn a large number of new, high-coverage English words in a fairly short time.

Is it possible to learn a large number of words in a short time? Probably not, if words can only be learned through random natural encounters in meaningful contexts. But research shows that this may not be the only way or even the best way of learning new words, particularly in a second language (L2). Numerous studies of L2 vocabulary acquisition have shown that contextual learning is slow (Horst, Cobb, & Meara, 1997; Mondria & Wit-de Boer, 1991), error-prone (Laufer, 2005), and poorly supported in frequency zones beyond the most basic 3,000 word families (Cobb, 2007). This is not to say that contextual encounters are not adequate for children learning their first languages over long periods, or that contextual information is not eventually needed by any learner to deepen the lexical knowledge once initiated by some other means. But for learners of either first or second languages who have somehow incurred a lexical deficit, it seems clear that waiting to meet words in context is neither a sufficient nor a necessary basis to catch up.

What is the alternative to contextual encounter as an input to word learning? Meara (e.g., 1995), Laufer (e.g., 2000), Nation (e.g., 2001), and Grabe (2009) have all argued in different ways for the need to build up a critical mass of vocabulary early in language learning, quickly, and out of context if necessary. The target for a critical mass has varied over the years but is currently placed at 5,000 to 8,000 word families, depending on the goals of the learner. The former figure will usually be adequate for conversational needs (Adolphs & Schmitt, 2003), the latter a starting point for academic language use including reading (Nation, 2006; Schmitt, Jiang, & Grabe, 2011).

While learning words out of context is clearly an initial type of learning, to be supplemented subsequently by rich contextual information, there is old research showing that such learning can be very fast and new research showing it can be effective in language learning. In a host of paired-associate memory studies from the late 19th century, subjects of normal intelligence were shown to be able to learn a vast number of L1-L2 word pairs or L2 word-meaning pairs (e.g., Ebbinghaus, 1885/1913) and to relearn them quickly after forgetting. This effect was further strengthened if the pairings were experienced in a pattern of recycling in which the time between pairings was systematically expanded (Gruneberg, Morris, & Sykes, 1978). The goal of this research was to study aspects of memory generally, rather than vocabulary acquisition specifically, but its bearing on acquisition has recently been taken up by Elgort.

In her doctoral study, Elgort (2007) gave advanced academic ESL learners practice in acquiring forty-eight English-like plausible non-words (PNWs, like *bance* or *benevolate*) using word cards with simple English definitions on the back. After only four hours’ practice with the cards over one week (recalling meanings from words, or vice-versa) in the spaced re-

hearsal rhythm mentioned above, the learners achieved native speaker levels of formal, semantic and procedural knowledge of the words studied including speed of lexical access in a range of priming conditions. Specifically, they had learned new words to criterion on Segalowitz' two indicators of native speaker automaticity, ballisticity (primed associations are unstoppable by conscious attention) and low variability in reaction times (Segalowitz & Hulstijn, 2005).

This extension of the paired-associate research augurs well for an approach to rapid vocabulary expansion. What Elgort's results do not provide, however, is a measure of use for the learned words, since it is not clear what PNWs can be used for. Also problematic is the pedagogical applicability of her finding. Whatever the evidence for the learning power of word cards (for a strong general argument see Nation, 2001, pp. 296-316), it is questionable whether any but the most determined learner would commit to a word card regime for enough words to make a difference. Further, while the PNW format obviated the need to decide on which real words the participants should learn, this question would feature in any real application. A more motivating way of realizing some of the benefits of matching specific words and definitions seems needed to extend Elgort's finding beyond the laboratory, and an extended game format could be one such way.

The state of play

A role for word games in vocabulary growth is hardly a novel idea although one that is not developed up to its potential in teaching practice. Meara (1995), a longtime proponent of building up a critical mass of vocabulary as early and quickly as possible in second language development, wondered why word games were not used more:

Word games do not provide the naturalistic, communicative contexts that language teachers usually think of when they are trying to provide contexts for using an L2. But, in fact, artificial contexts of this sort provide a very good environment for using words. [...] Word puzzles are incredibly popular with L1 speakers, and it is surprising that language teachers have not exploited this popularity more. (p. 11)

Indeed, the many word games available via popular culture, newspapers or Internet depend on the very word learning operations indicated over and over again in acquisition research. Retrieval of form from meaning and meaning from form, word recognition, and lexical holding capacity are all involved to varying degrees in Scrabble, acrostics, crossword puzzles, word jumbles, and the like. Further, the case for learning from games is presumably made stronger in the meantime with the added possibilities on both gaming and learning fronts from the integration of games with electronic and computational media. The under-use or peripheral use of games in language learning is therefore hard to understand, and almost certainly reflects a prejudice among practitioners, course writers, and maybe learners themselves against both learning out of context and learning from games. But as research raises the respectability of both of these activities, practice is likely to follow.

Practice has already begun to follow in some respects. Vocabulary games are popular on the Web (e.g., the UN Food Program's *Free Rice* at <http://www.freerice.com/>), and numerous vocabulary learning applications are now available for the iPhone (Godwin-Jones, 2010). Nakata (2011) reviews a number of electronic word card learning systems available on the Web, noting, however, that the design principles, learning outcomes, and even exploitation of existing technologies in these systems tend as yet to be rudimentary. Indeed the investigation of learning through gaming remains weak in general, being described by the editors of a recent gaming issue of *Educational Technology Research and Development* as being "in its infancy" (Spector & Ross, 2008, p. 510). To conclude, the stage is set for gaming in lan-

guage learning, particularly in the case of vocabulary acquisition, but serious play is yet to begin.

RESOURCES AVAILABLE FOR A VOCABULARY TRAINER

The idea that a computer-based training system could be a good way to meet the vocabulary challenge outlined above has been around for some time. The large amount of material to be covered, the likelihood of strong individual differences in both goals and learning rates, and the need for recycling and record keeping, were early identified as reasons to look to computerized instruction (e.g., Atkinson, 1972). Recent developments merely add to the argument: increased capacity allows for the provision of speech, concordances, or glossaries as learning tools; increased processing speed allows for fine-tuned control of procedural interactions; and the advent of networks frees learning from the particularities of time or place. The network advantage grows stronger almost daily as game players like Nintendo assume network capabilities, and smart mobile devices like the Blackberry and iPhone become in effect portable computers.

A less obvious link between computing and vocabulary learning is that computer-based lexical analysis, particularly but not only of English, is the basis of much of our current understanding of vocabulary acquisition. Frequency analysis of large corpora has allowed the prioritization of specific learning goals; collocation analysis has revealed the extent of lexical patterning and repetition in language use; reaction time instruments have revealed the role of word recognition and automaticity in language processing; and corpus-based lexicography has created a revolution in the quality of dictionary resources dedicated to language learning. Admittedly, these computer-based tools and insights can be exploited pedagogically in any medium, but the computer medium is a natural choice for the job (Cobb, 2008). The computational capacities deployed hitherto analytically are also useful instructionally, for example, the capacity to store, sequence, and recycle large quantities of language, to track and control reaction times, and many others.

From the vocabulary research and development since roughly 2000, here is one highly realizable set of desiderata for a suite of vocabulary games.

Syllabus

Only recently has it become possible to specify the complete basic or non-specialist lexicon of English such as could form the syllabus of a comprehensive vocabulary trainer. This specification follows from four connected computational research projects over a ten-year period: the assembly of the 100 million word British National Corpus (Oxford Computing Services, 2000); its breakdown into a list organized by frequency and range (Leech, Rayson, & Wilson, 2001); the pedagogical adaptation of these lists including family groupings (Nation & Beglar, 2007); and their exploitation of these lists in coverage studies for particular needs and frequency zones (Nation, 2006).

Dictionary

Also relatively recent is the publication of electronic learner dictionaries with simplified definitions, following the principle that as few words as possible in the definition should be less common than the headword itself (as was always the case with the format "a car is a vehicle which..."). The *Cambridge advanced learners' dictionary* (2005) is arguably the current leader in this field in English, owing to its balance of brief comprehensible definitions and adequate number of entries. A question arises as to whether the definitions are most usefully presented in the learners' L1 or L2. Elgort (2007), for example, chose target-language definitions (English words and English definitions for ESL learners) following research by Jiang (2004). Jiang's semantic transfer model of L2 vocabulary acquisition postulates that

new L2 words are almost inevitably associated with old L1 concepts for extensive periods, or forever, if learned through L1 definitions, unless steps are taken to prevent such associations from forming. Elgort found this could be achieved through pairing L2 words and L2 definitions through several rehearsals, provided of course the definitions are short and comprehensible. She attributes the results of her study to this specific factor.

Testing

A testing format for placement and achievement tracking that is compatible with a game format is the simple Yes-No checklist test based on signal detection theory and developed for L2 vocabulary measurement by Meara and colleagues. Such a test simply asks learners whether they know each word on a particular frequency list, yes or no, and relies on plausible non-word (PNW) items in the list to keep a check on honesty. Meara and Buxton (1987) developed the algorithms to modify scores according to the number of PNW choices and tested the test's predictions with large numbers of learners in medium-stakes settings (a collection of such tests can be found at <http://www.lognostics.co.uk/>, click 'Tools').

Recycling

It is well known that words have to be encountered several times in order to be retained (see Zahar, Cobb, & Spada (2001) for a discussion of the number of times), but in a game context where motivation and variety are priorities the question is rather how few times will suffice. Mondria and Mondria-De Vries (1994) propose a regime based on the paired-associate research already mentioned for their "hand-held computer," which is basically a shoe-box with five compartments of increasing size holding word cards with new words on one side and short definitions on the other. This simple technology attempts to exploit the classic finding that paired-associate learning is maximally effective if associations are reviewed just before they are forgotten, and that such reviews should occur in a "spaced distribution" (since forgetting typically takes two to three times longer to occur after each review). New words move through the shoe-box and are reviewed, word to meaning or meaning to word, and promoted or demoted from slot to slot. As both new and more difficult/less frequent words are added to the first for more frequent review, promoted words are reviewed less and less frequently, eventually departing the game entirely via the fifth compartment.

Focus on form

The main learning operation proposed thus far is form-meaning mapping. In fact there is also an argument for including games that focus on word form alone, or on form and meaning separately. Indeed this is a key recommendation from the input processing research (e.g., VanPatten, 1990), with its concern for information overload in early or pre-automatized language learning. The specific value of giving independent attention to form is based on research showing that the form part of form-meaning connections are often weak in naturalistic vocabulary acquisition; the emphasis on meaning in contextual inferencing often leads to global comprehension but no retention for the novel word form itself (e.g., Mondria & Wit-de Boer, 1991). Indeed, the establishment of a form in the mental lexicon is likely to be a much slower process than putting together a meaning (which in any case is available through both general knowledge and of course the L1 lexicon, as suggested in Ji-ang, 2004) as well as being less amenable to explicit learning (Ellis, 1994; Hulstijn, 2002).

Focus on both receptive and productive learning

Different games or parts of the same game should ask players to both recognize words qua forms, recognize meanings for these words, and produce words from meanings. Meaning recall is unfortunately not simple to operationalize in a game format (show that you know

what *cat* means), but production and partial production of words in response to meanings can be accomplished in menu choices or spelling tasks.

Focus on lexical access

Lexical fluency, as normally represented by the number of milliseconds needed to make lexical decisions about single words out of context, is one of the strongest predictors of both degree of consolidation of word knowledge and successful reading comprehension, as shown in both L1 and L2 studies (Grabe, 2009, Ch. 2). The automatization of lower level lexical access processes frees up the memory resources needed for processing higher level meaning and novelty (Segalowitz & Hulstijn, 2005). It is often argued that lexical access speed, like other implicit or procedural knowledge forms, can only be built up over thousands of hours of exposure to a language (e.g., Ellis, 1994); but some recent laboratory studies by Snellings, van Gelderen and de Glopper (2002) and Elgort (2007) give hope that lexical access is to some extent trainable.

GAME DESCRIPTION

These resources and opportunities for vocabulary gaming are realized in *Word Coach* as follows. The learning content of *Word Coach* is effectively the entire contemporary, non-specialist lexicon of the English language, as represented by its 14,000 most frequent word families, linked to a specially adapted version of the *Cambridge advanced learner's dictionary* (2005) purchased by the gaming company from Cambridge University Press specially for this project. The definitions were shortened and in some cases polysemous senses combined for a more general meaning by one of the authors' research assistants. While it is debatable whether the full lexicon should be the object of direct instruction, particularly in a second language (Nation, 1990, for instance, argues for the direct instruction of only the 3,000 most frequent word families followed by refocusing on contextual inference strategies in view of the declining coverage of the remainder), recent coverage studies, as already noted, have raised the direct learning figure to more like 8,000 families, and in any case, the goal of meeting the needs of both L1 and L2 learners requires the inclusion of L1-oriented targets.

The fourteen content levels are broken into 1,000-family sections, one of which defines a learner's zone of play at a given point in time. Learners are tested at the beginning using an adapted version of the Yes-No checklist format described above (and depicted in Figure 1, left panel), assigned a frequency zone to work in, and then begin playing the simpler of the several word games with randomly selected words from this zone (described below). All words that pass through the game are recycled at least five times. Words that a player makes no errors with pass through the game in just these five encounters, with increasing space between appearances; words that cause a player to make an error, whether in respect to form or meaning, are scheduled for more extensive recycling according to the algorithm proposed by Mondria and Mondria-de Vries (1994, as depicted in an animated figure on the game's Website at http://mywordcoach.us.ubi.com/expression_potential_how_works.php under "the five box rule").

Word Coach's current lexical inventory for a given player consists, like Mondria and Mondria-de Vries's (1994) shoe box, of five compartments. Words begin their game-life in the first compartment, and if played correctly, move to the second, and so on until they exit the game in a minimum of five correct plays. If at any point they are played erroneously they are demoted back to the previous compartment. The effect of this to-and-fro of words is that as play proceeds there are many words in compartments four and five, which are being recycled in a game only rarely and at increasing intervals because they do not come up very often, and a moderate number of words in compartments two and three, a mixture of novel

and repeated items, which are being recycled more frequently. The number and difficulty of both words and games evolves as play proceeds. The game keeps detailed records, provides regular feedback, and moves the player to higher levels (to meet lower frequency words) as learning criteria are met. The games focus on one or a mixture of the following skill areas: form, form-meaning connection, and lexical access.

The user loads *Word Coach* into the Nintendo console, enters his or her name (four can play on one machine, plus guests, and many more wirelessly) and begins to play the one game available at this point, which is effectively a placement test (a sample item is shown in the leftmost screen of Figure 1). The testing starts by pitching word sets from a medium frequency level (the 4,000 word level); if these words are not known, the level goes down toward more common words (or if known goes up toward less common words) until a zone with more unknown than known words is found (at least 50% unknown), and this is where play begins. In this way, most players should be challenged at their own level at the beginning of play. A point to note about the testing is its surface similarity to, but fundamental differences from, the testing literature on which it is based: far more words are sampled to determine a level in the testing literature, and the criterion for mastery of a level is more like 80% than *Word Coach's* implicit 50% (e.g., Nation, 1990). The goal of these modifications is to balance correct placement with a quick and motivating or game-like test.

More games and variants of games are introduced as play proceeds. There are two form-based games; *Missing Letter*, which involves using the Nintendo's stylus to supply the letter missing from a word (e.g., *new_paper*, Figure 2, left screen), and *Block Letters*, which involves clicking on falling letters to form one of four given words, as unused letters pile up toward an explosion (a version of *Tetris*, Figure 2, middle screen). Four of the games involve connecting words and simplified L2 meanings from the modified Cambridge lexicon, in various combinations of both receptive and productive tasks. In *Split Decision*, the upper screen displays a definition while the lower screen presents words the players must toggle through until they identify the one that matches the definition. In *Word Shuffle*, a word from the bottom screen is dragged and dropped on one of four definitions in the top screen. *Pasta Letters* shows players a definition whose corresponding word they must produce by dragging its letters in sequence out of a bowl of alphabet soup before they sink out of sight (shaking the bowl can raise the letters back to the surface for a time). *Safecracker* presents a definition whose word players must spell on the dial of a safe but now in competition against an opponent, either human or machine generated if no human is available. The games are played in sets of twenty words and advance from basic to more challenging versions. The intermediate version of *Missing Letter* challenges the player to identify the wrong letter from *newpaper* and correct it, rather than simply supplying a missing letter. Then in the advanced version, correct words that require no change are added to the mix. Lexical access speed is emphasized throughout, by advancing the rate at which the words drop, at which the electronic opponent drags his letters out of the soup, and so on, and of course by levying rewards and penalties in the post-mortems that follow each game set.

At the end of each set, missed words are reviewed, points tallied (with bonuses and penalties according to time taken), elaborate progress graphs presented, errors highlighted, and meanings reviewed (see Figure 2, middle). One of four focus group-created cartoon tutors review speed, errors, and persistence, offering encouragement, advice and mild admonition ("You've been absent for a long time"). Players are recommended to work through 100 words per day, and at intervals they assume new levels of "expression potential," each level being named for the language ability loosely associated with a lawyer, a reporter, a poet, and so on. This dubious but probably harmless idea, invented at the commercial end of the design process, addresses the legitimate concern that 1,000-levels were both boringly named and too big to pass through in a reasonable amount of time, entailing excessive de-

lay between promotions. Figure 1 shows the placement-test format on the left; the other screens show a selection of the management tools which encourage players to reflect on their learning; Figure 2 shows a sample of form-based game screenshots, the deployment of the split screen, and one of the several types of error feedback; and Figure 3 shows a sample of form-meaning games (for the others, see the game’s Web site at <http://mywordcoach.us.ubi.com/>).

The games and sequences are thus an attempt to exploit the resources and realize the design principles outlined above. As such, do they lead to any measurable learning in the three areas they target – form retention, form-meaning mapping, and lexical access?

Figure 1
Placement test and learning management tools

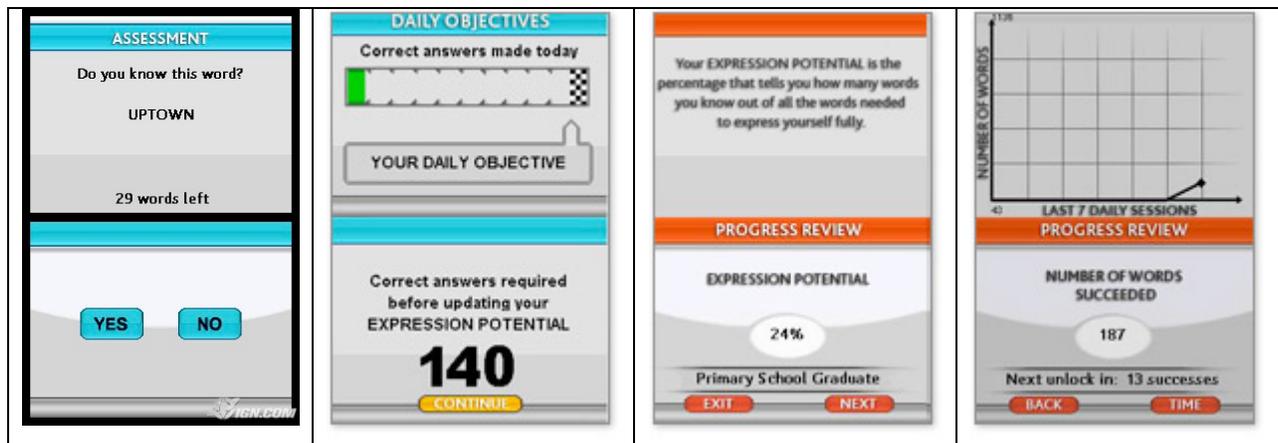


Figure 2
Two form-based games and session feedback



Figure 3
Three form and meaning games



METHODOLOGY AND RESEARCH QUESTIONS

The goals of *Word Coach* are to present learners with a principled input of new words with comprehensible definitions in their vocabulary growth area and recycle these words via games that emphasize different aspects of word learning and are fun to play. The design is meant to ensure that as many as possible of the words presented will be remembered, understood, processed fluently, and eventually used. There are many potential research questions within this agenda, but in this initial study of vocabulary learning by video game the following questions are asked:

1. What is the extent and nature of game use, and are there any game or learner characteristics that predict game use?
2. How many word meanings are known receptively before and after game use?
3. Is there a difference in speed of lexical access before and after game use?
4. How many learned words are used in free production after game use?

A backdrop to these explicit research questions is a further implicit question about the degree of trade-off between learning principles and the mass playership accessed via commercial gaming.

Learners and setting

The learners who participated in this study were two intact classes of twenty-five grade six Francophone ESL learners in a middle-class suburban school in Montreal (eleven to twelve years old). The ethnicity of both groups was roughly 30% Quebec Francophone children and 70% immigrant children from mainly Francophone countries. The medium of instruction at the school was French, except for two hours per week of ESL, in which the teacher used and asked for English as much as possible following a communicative teaching methodology. The children had various amounts of exposure to English out of class, from extensive to none, and widely varying levels of English proficiency. The ground had already been prepared for the topic of learning through games inasmuch as parents had complained that non-educational video games were eating into their children's homework time, and 98% stated they were more than willing to try the educational variety. The school supported the research, which took place over four months in the spring of 2007. Ubisoft Canada of Mont-

real provided the loan of thirty Nintendo DS Lite players and *My Word Coach* game disks, enough for one class at a time to use the game, and no restrictions on their use. The school stipulated that all children involved in the study should have equal access to the game over the four months.

Research Design

Because of the school's requirements that all children have a chance to use the game, and that the research groups be intact classes, a quasi-experimental, within-subjects design was chosen for the study. One group used the game for two months while the other served as quasi-control, and then the roles were reversed. On receiving the game, each group received fifty minutes of training and encouragement to use the game from the teacher. The same series of word knowledge tests were administered to both groups at the beginning of the experiment, then at the two-month changeover point, and then again at the end of the four months. For learners who received the game first, the sequence was as follows: pre-test, two months using the game, post-test 1, two months of normal classes without the game, and post-test 2. (The second post-test thus served as a measure of delayed retention for one group.) For learners who received the game second, the sequence was pre-test 1, two months without the game, pre-test 2, two months with the game, and post-test. (The two pre-tests thus served as a baseline for one group on normal lexical development against which game related development could be compared.) These groups are referred to as Group 1 and Group 2 in the remainder of the study, and the three testing points as T1, T2, and T3 (the design is shown schematically in Table 1).

Table 1
Diagram of research design

	T1		T2		T3
Group 1	Pre-test	2 months classroom + game use	Post-test 1	2 months classroom	Post-test 2
Group 2	Pre-test 1	2 months classroom	Pre-test 2	2 months classroom + game use	Post-test

Experimental conditions were far from laboratory like. The two intact classes were present in the same school, some students were known to possess their own Nintendo players, *Word Coach* was available commercially at the time, and no attempt was made to assure that learners had no contact with the game in their non-game periods (indeed the game is set up to encourage multiple players). However, the teacher who was in charge of both groups reported never observing a game player in the hands of a learner outside the game period, although of course the students were in free possession of the materials for two months and there is no way of accounting for evenings and weekends. All measures were compared at the three test points using basic one-way ANOVAs followed by Tukey post-hoc tests; the usual arrivals, departures and absences of an intact setting while not extensive created slightly unequal groups making a repeated measures study inappropriate.

Research Instruments

The research instruments were as follows:

1. The game itself provides detailed tracking of which words were played and how often, as well as session size and frequency (Figure 1).

2. The recognition knowledge measure is Nation and Beglar's (2007) Vocabulary Size Test, a revision of the classic Vocabulary Levels Test (Nation, 1990). The levels for the Size Test are sampled from the first fourteen BNC frequency lists, as elaborated by Nation, and thus correspond precisely to game content. The test measures only recognition knowledge, in that the test taker is not asked to produce the word or its meaning but merely to match test words in short non-defining contexts to one of five glosses (shown in Figure 4). The test glosses are not the same as the glosses encountered playing *Word Coach*, except coincidentally.

Figure 4
Vocabulary Size Test (2007) example

<p>PERIOD: It was a difficult period.</p> <p>a. question</p> <p>b. time</p> <p>c. thing to do</p> <p>d. book</p>

The Size Test has ten questions at each 1,000-family level, such that the score multiplied by 100 gives an estimate of the number of word families known at that level (8 out of 10 = 800 families known). Nation and Beglar (2007) discuss the test's sampling and reliability. In this experiment only the first ten levels of the test were administered, in view of the learners' predicted level and the institution's time constraints. Since only one validated version of the test was available at the time of the experiment, the same test was used with each learner three times. It should be noted that the tested words were not necessarily ever encountered in the game, but are drawn from the same frequency zones the learners were playing at. A test based on the exact words played would have been difficult to construct (given the wide disparity of learner abilities, the large number of words involved, and the need for a modestly sized test that could be completed in under fifteen minutes) and would not have provided a pre-test measure.

3. Lexical access speed was measured with a simple instrument developed by UNESCO for literacy testing in developing countries. The instrument is simply a paper list of sixty high frequency words (from the first 1,000 frequency zone) arranged in order of decreasing frequency and increasing length. Learners are asked to read as much of the list aloud as they can in one minute; a research assistant notes the last word reached within the time limit and strikes out any mispronunciations that appears to indicate unfamiliarity with the word, resulting in a tally of number of words read correctly in one minute.
4. The production measure was an oral description of the twenty-five line drawings of Mayer's (1967) wordless story *A Boy, a Dog, and a Frog*, as told (untimed) to a research assistant and recorded. This took place immediately after the lexical access trial. The recordings were later transcribed as text files for processing by the BNC version of *Vocabprofile* (at <http://www.lexutor.ca/vp/bnc/>). This text analysis routine categorizes each word of an input text by 1,000 levels, according to the same familiarized word frequency lists and vocabulary size test described above, resulting in a frequency profile of the learner's production (the percentage of word families, types, and tokens at each 1,000 level). Any L1 items were

excluded from the analysis. The rationale for getting this information was that it would enable a comparison at each level and between receptive gains and productive use.

All four measures were administered in a single hour in each of the three testing periods. In addition, the teacher solicited written comments about their game experience from the learners using an informal questionnaire format.

RESULTS

Game use

According to *Word Coach's* tally of "words succeeded" (in the game's vernacular) the amount of game use in each of the two-month periods was similar between the groups but with considerable variance among individuals. Group 1 players succeeded an average of 2,849 words (SD=1,879), while Group 2 succeeded 2,536 words (SD=1,959). The difference between means is not statistically significant, and the standard deviations are high showing that some took to the game more than others (the range for words succeeded is 6,400 down to 272). These numbers translate into days and hours as follows: At the high end, there were 6 players in each group who succeeded at over 4,000 words in 60 days, which at an estimated average of 8 appearances of each word needed to remove it from the game, totals roughly 32,000 words played, or just over 500 words per day, about 25 game sets of 20 words apiece. At 5 minutes per game this extrapolates to about 2 hours of play per day, or a total of 120 hours over 2 months. At the low end, there were 7 in each group who succeeded at fewer than 1,000 words (8,000 recurrences, 134 words per day, 6 or 7 game sets). At 5 minutes per game, this amounts to about 30 minutes per day, or a total of 30 hours over 2 months.

As a further investigation into any common characteristics behind high and low game use, two usage groups of equal size were formed from the total population by rank ordering number of words succeeded and then dividing these into a high use group (those who had succeeded in more than 2,000 game words, 6,403 down to 2,080), and those who had succeeded in fewer than 2,000 (1,968 down to 450). Usage group membership was then compared to performance on the different levels of the Vocabulary Size Test at T1. The interesting relationship was between words succeeded and scores at the first 1,000 level. Of the heavy users, 70% had 1,000-Level test scores of 60% to 70%, suggesting they probably knew about 600 or 700 of the first thousand words of English; of the light users, 75% had scores that were either below 400 words or above 800 words. Whether any of this information allows a prediction of game use will be explored in the Discussion.

Receptive word meanings known

Both groups took the first ten levels of the Vocabulary Size Test of meaning recognition (100 items) at all three testing points. It does not appear that the learners had "learned the test" to any large degree since pre-post scores for Group 2 when they did not use the game are not significantly different (see Table 3).

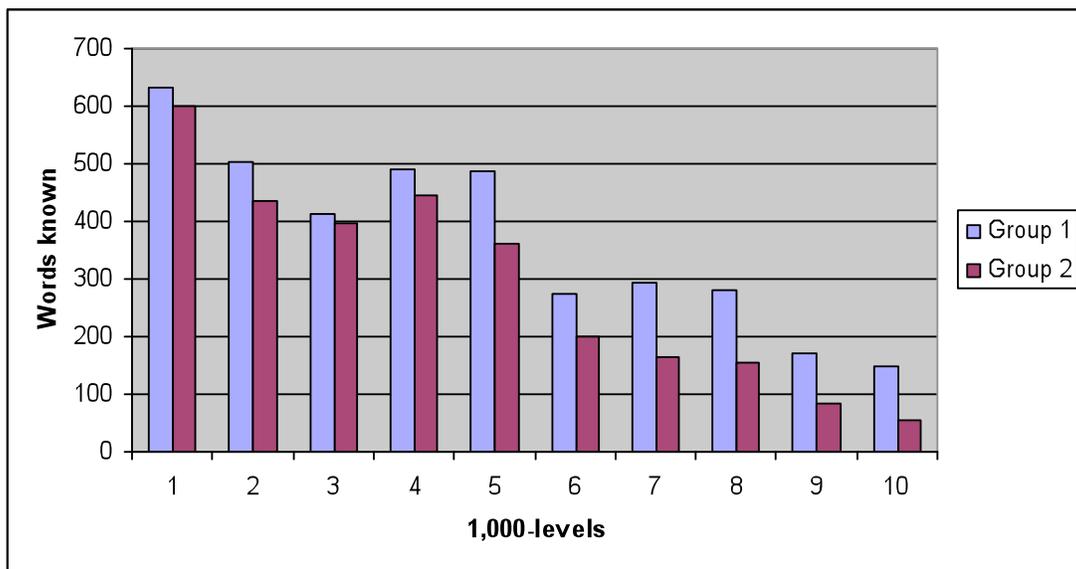
The pre-test results at T1 were similar between groups (shown in Figure 5 and Table 2), with the appearance of a slight advantage for the first game group (not significant for any single level below the fifth, nor for the first five taken together). A point to note is the roughly equal numbers of words known across the second through fifth levels, between 40% and 50%. The scores drop sharply after the fifth level, however, and because the first 5,000 words are a sufficient immediate goal for these learners, learning gains were calculated for only the first 5,000 word families. The mean number of words known from the first 5,000 at T1, as calculated by multiplying mean level scores by 1,000 and adding the totals

in the procedure outlined above, was 2,428 (SD=771) for the first group, and 2,236 (SD=873) for the second, leaving a comfortable learning space of at least 2,500 words.

Table 2
Mean Levels Test scores for first 10,000 word families at T1 (SD)

	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
Group 1	6.33 (1.51)	5.03 (2.39)	4.13 (2.17)	4.90 (1.64)	4.87 (1.93)	2.73 (2.06)	2.93 (1.57)	2.80 (2.07)	1.70 (1.53)	1.47 (1.67)
Group 2	6.00 (2.04)	4.36 (2.53)	3.96 (2.46)	4.44 (1.86)	3.60 (2.64)	2.00 (1.94)	1.64 (1.49)	1.56 (1.72)	0.84 (1.05)	0.56 (0.90)

Figure 5
Word families known at T1



Following game play, learners in both groups acquired new words across the first five levels of the test fairly evenly, as shown in Figure 6 for the first group (the group for which pre-, post- and delayed post-test scores are available in the present design). Table 3 shows the number of words learned across five levels by both groups.

Figure 6
Mean Levels Test scores by level for Group 1 at three times

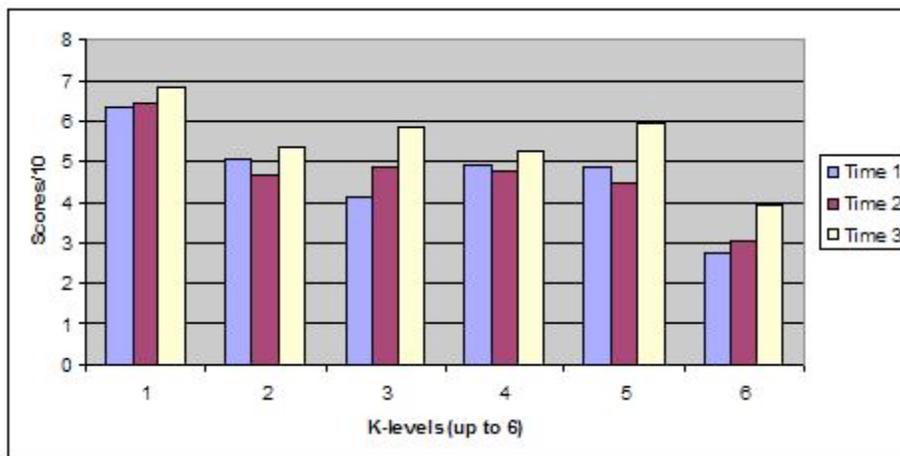


Table 3
Word families from 5,000 known at three times +gain, (SD)

	Time 1	Time 2	Time 3
First game group	<u>2428</u> (771)	<u>2527</u> (739) +99	2932** (702) +405
Second game group	2236 (873)	<u>2332</u> (779) +29	<u>2556*</u> (838) +224

** p<.001; *p<.01; underlining indicates game periods

The results at T3 show significant and extensive growth of word knowledge. For the first game group, the average number of meanings recognized had increased by only 99 immediately after the game (n.s.d.), but then at a delay of two months increased by a further 405 words ($F=9.19, p<.001$), for a total of 504 words over the four-month period, or a gain of 20.7% with respect to the starting knowledge of 2,428 words. For the second group, the average number of meanings recognized had increased by only 29 words through two months' of normal classroom experience (n.s.d.), but following the game had increased by 224 words (significant over T1-T3, $F=5.11, p<.01$), a gain of just over 10% with respect to the mean starting knowledge of 2,236 average words.

Increased speed of lexical access

Both groups made small gains in lexical access response times in the non-game periods (presumably a practice effect for the testing method) but significantly larger gains in game periods. From a similar starting point of 54.9 average words read correctly in one minute for Group 1, and 63.6 words for Group 2 (n.s.d.), Group 1 increased their number by 19.5 words or 36% from T1-T2 and another 6.9 by T3 ($F=20.82, p<.0001$). Group 2 increased its words read by an average 8.2 words at T1-T2, and by 12.2 more words or 18% increase following the game period T2-T3 ($F=4.29, p<.05$, see Table 4).

Table 4
Number of error-free high frequency words read in 1 minute at three times (SD)(+gain)

	T1	T2	T3
Group 1	<u>54.9</u> (11.4)	<u>74.5**</u> (27.9) +19.5	81.4* (30.4) +6.9
Group 2	63.6 (20.4)	<u>71.8</u> (25.2) +8.2	<u>84.0*</u> (28.2) +12.2

** p<.01 *p<.05; underlining indicates game period

New words in use in the Frog Stories

The Frog Story accounts were transcribed by research assistants, spell-checked manually, stripped of immediately repeated words and phrases by a computer program, and run through *Vocabprofile*-BNC both as individual tellings and as a combined group corpus. Only the group corpus results are reported here. Since the vast majority of played words were at the 2,000-5,000 level, as well as the gains in receptive knowledge, as shown above, a way of determining whether newly learned words went into active use was to check the story corpus for 2,000-5,000 items. The specific question for each corpus then was how many of the second, third, fourth, and fifth 1,000 level words that had not been present in the first telling would show up after game use. The simple answer was that almost none showed up. The VP-BNC analysis revealed that the stories were composed overwhelmingly of first-1,000 level words, apart from the 4,000-level word *frog* and one or two others which the learners already knew (there was a labeled picture of a frog in the English classroom), and that the pattern did not change across the period of the experiment. Results for Group 1 are shown in Figure 7 (results were virtually identical for Group 2).

Figure 7
Profiles of the frog story corpus

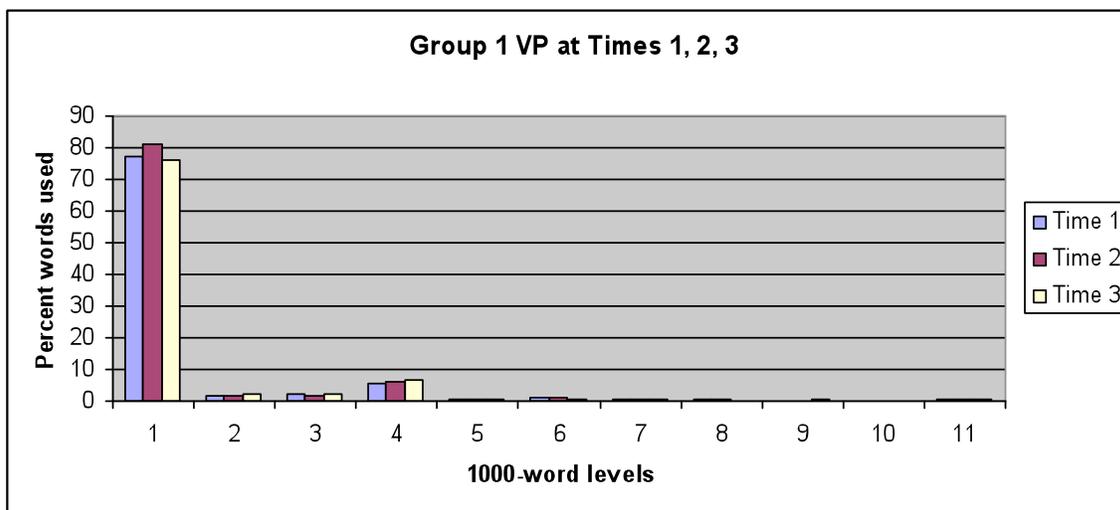


Table 5 shows *Vocabprofile*'s family, type, and token counts for the two groups at three times. There are an average three additional word families for each group following the game periods (at T2 for Group 1, T3 for Group 2), but these gains are not statistically sig-

nificant ($p < .10$ in both cases). There were, however, increases in the number of total words (tokens) used to tell the stories following game use, at T2 for Group 1 (40 mean increase from an original mean of 167, about 25%, $F = 5.78$, $p < .01$) and T3 for Group 2 (20 mean increase from an original 159 at T2, about 12% but not significant). There are declines for all units in repeated re-tellings following non-game periods, whether preceding game use (Group 2 at T2) or following (Group 1 at T3).

Table 5
Frog Story mean word counts (SD), +/- gain

	Time 1			Time 2			Time 3		
	Fam	Typ	Tok	Fam	Typ	Tok	Fam	Typ	Tok
Group 1	<u>45</u> (22)	<u>51</u> (26)	<u>167</u> (87)	<u>48</u> (21) +3	<u>54</u> (26) +3	<u>207**</u> (90) +40	44 (21) -4	50 (26) -4	179 (75) -15
Group 2	40 (13)	45 (14)	157 (42)	<u>38</u> (16) -2	<u>41</u> (19) -4	<u>159</u> (52) +2	<u>44</u> (17) +3	<u>48</u> (21) +2	<u>179</u> (60) +11

** $p < .01$; off-list items are not included

Student questionnaires

After each game period, the teacher took semi-structured written reports from the students on their experience of the game. Most students expressed satisfaction with their experience, seeming to prefer the games with less to read (single words rather than definitions) and faster action (like *Block Letters*). One recurring comment stands out, that the players perceived the systematic reappearance of words as "boring," whether new or previously involved in an error.

DISCUSSION

A pattern to notice across the measures is that while results are in the same and largely positive direction for the two groups following the game periods, they are larger and more consistently significant for the first group to use the game, despite the two groups' equal starting points and extent of game use. It is probably safe to say that there was an excitement to the first run of the game that had dissipated slightly when the game had become habitual. Another pattern is that on two measures the gain or rate of gain drops off right after game use for at least one group (Group 1's story size and lexical access), which when joined with the previous pattern may suggest that a longer period of game play is needed to consolidate learning, yet that the motivation needed to sustain a longer period could not be taken for granted (7 out of 25 in each group or 28% did not appear to take to the game, and the meticulous recycling of items was perceived by some who did as "boring.") A third general pattern is that surprise findings were generated on a number of measures, including that the largest meaning recognition gains were registered two months rather than immediately after game play, and that the apparent effect on production was not the appearance of newly learned words but improved access to words almost certainly known already. Both surprises reinforce the idea mentioned in the introduction, that games-based learning research is in its infancy (perhaps it is a little less so now). Other patterns pertain to the individual game objectives.

Words learned

Based on the performance of the two groups of learners, it appears that an average of between 224 and 504 of the words presumably met largely on *Word Coach* were remembered

and understood to the extent that they could be transferred to a matching definition that was not simply the one originally learned in the game. Given initial mean receptive vocabulary sizes of 2,428 and 2,236 in the 5,000-word zone, this is an expansion of just under 10% (Group 1) and just over 20% (Group 2) respectively. The lower figure could plausibly turn out to be higher if it had been possible to retest Group 2 two months after game play, since it was in the post-game period that Group 1 gains were apparently consolidated.

The importance of these gains should be assessed against two backdrops, the mean numbers of game words that had been through a full recycling ("succeeded"), and the normal vocabulary growth that occurs in classroom learning. The mean number of words that had been played successfully at least five times and hence succeeded was 2,536 and 2,849 words for the respective groups, such that 504 words or just under 20% of succeeded words were remembered by the first group, and 224 words or just under 10% were remembered for the second group. The similarity of vocabulary size percentage gains and succeeded words retained is based on the coincidence that T1 vocabulary size and succeeded words were similar (2,428 words known to Group 1 at T1 and 2,849 words succeeded; 2,236 words known to Group 2 at T1 and 2,536 words succeeded).

As for the typical or habitual vocabulary growth for ESL learners in Quebec schools, this has never been directly measured, so Group 2's 29 average new words T1-T2 is probably a good start toward building a baseline. Such a figure is plausible, in that 29 words in two months extrapolates to 145 for a 10-month school year and about 1,160 for eight years of school ESL, with some multiplier differential coming into effect as proportion of words known gives higher text coverage (discussed in Zahar, Cobb, & Spada, 2001). This figure tallies with both international norms (Milton & Meara, 1995¹) and studies of comparable but older Montreal learners (Zahar et al., 2001).

As already discussed, the strongest knowledge gains (Group 1's 405 words T2-T3) were registered not immediately following the game period, but at a two-month delay. In fact, delayed appearance of lexical growth is not unprecedented (it is also found in Cobb, 1999), and may even be the norm, unacknowledged owing to the general lack of delayed post-test measures in acquisition studies. Words learned inside one's growth area are often encountered again within a short period, so that further consolidation takes place, in an "as soon as I learned that word I saw it everywhere" phenomenon.

A further point of interest in these results, and one with bearing on broader questions of learning and commercial gaming, is the frequency distribution of the words learned. As mentioned in the results, the Size Test showed similar percentages of words known across the second through fifth frequency levels (between 40% and 50%) at both pre-test and post-test. Indeed gains were if anything stronger at the third and fifth thousand levels than the first (as suggested in Figure 6). This spread across the frequency bands almost certainly reflects the testing procedure adopted, of working down or up from the four thousand level until a level with roughly 50% unknown was determined. In this case, this would have happened at third thousand level for most of these learners, on the assumption this was the highest level unknown, whereas in fact there were often two or three 1,000 levels with 50% or more unknown below that. Most of these learners would have ended up playing at the first 1,000 level, their true growth area, if the research-indicated 80% criterion had been adopted instead of the game-inspired 50%. In effect, the game as constituted assumes that the frequency levels are learned in sequence, as seems largely true in L1 (Biemiller & Slonin, 2001) but for various reasons may not be in L2 (Cobb, 2010). Given the coverage importance of the first two 1,000 levels of English (at least 85% of lexical items in most of what these learners would read), knowing only 50%-70% of the items in this zone would seem to be a disadvantage, and one not greatly affected by all the words learned at other

levels on *Word Coach*, owing to a testing system that inadvertently lumped together L1 and L2 learning needs and prioritized gaming over learning.

Access speed

The increase of between 18% and 30% in the number of words that could be recognized and decoded in the space of a minute adds to the ongoing conversation about whether or not lexical retrieval is “an aspect of fluent L2 production that can be enhanced” (Snellings, van Gelderen, & de Glopper, 2002). The emerging consensus is that it can, and the results presented here are a further positive demonstration. The next question is whether training gains will transfer to resource sensitive operations like reading, and this is presently unclear (Fukkink, Hulstijn, & Simis, 2005). One thing that seems clear is that fluent lexical access in an L2 does not develop by itself. Segalowitz and Hulstijn (2005) found that ostensible bilinguals in Montreal often had slower and more variant lexical access times for common words than native speakers, suggesting they used attention-demanding resources rather than automatized lexical processing for reading. Speeding up is not the same as automaticity, but it is presumably a step along the way, and Elgort’s (2007) strong finding for a similar training regime suggests that gaming and lexical access is a relationship worth developing. Faster information processing and reaction times are probably a principle advantage of this particular learning technology (as shown often in studies of flight simulators etc., and now for lexical access). As mentioned, these gains in lexical access were for common or high frequency words, words probably already known to most of the participants, rather than for the less frequent words actually learned in the game. While such a result runs counter to common sense, similar phenomena have been observed in other language acquisition studies (e.g., Laufer & Nation, 2001) and are under current investigation. A plausible mechanism may be that an influx of new items improves access for already known items.

Words in use

Oral production results show that while few if any entirely new word families went into immediate use in the Frog Stories, the stories nonetheless changed on other measures. Story size grew significantly following game use, leading to the question of where these extra words came from. The presence of extra word tokens indicates some additional elaboration of incidents in the story (not just repetitions, because these had been removed). A post hoc investigation of the Frog Story *Vocabprofiles* revealed that the main source of these extra English words was a dramatic reduction in the number of French words, i.e. in code switching. The off-list component of the original *Vocabprofile* outputs was almost exclusively L1 items, which, as already mentioned had not been counted in the story size tallies. This component dropped by 52% for Group 1 ($F=8.97$, $p<.01$ at T1-T2, and $p<.05$ at T1-T3) and 40% for Group 2 at T2-T3 ($p=.06$, not significant, see Table 6). It appears that English words had become available for telling the stories in places where, prior to game use, only a French word had been available. These additional English words were presumably not newly learned words, since the number of new families produced did not grow substantially for any zone, but better access to previously known words.

Table 6
Percentage of French tokens in Frog Stories at three times (SD)

	Time 1	Time 2	Time 3
First game group	<u>10.3</u> (11.7)	<u>5.4**</u> (8.3)	7.2* (10.4)
Second game group	12.0 (10)	<u>10.0</u> (9)	<u>6.0•</u> (7)

** p<.01; *p<.05; •p<.06

Predictors of game use

As mentioned, the investigation of usage patterns showed the population breaking down into equal high and low usage groups, with 70% of heavy usage groups knowing 600 to 700 words at the 1,000 level of the Vocabulary Size Test, and 75% of light users knowing either below 400 words or above 800 words. Can any sense be made of this pattern? It appears that too much or too little knowledge of the first 1,000 words predicted light game use, while moderate knowledge predicted heavy game use. Given the crucial role of the most frequent 1,000 word families in English at large (comprising at least 70% of an average text and far more of an average conversation) and *Word Coach* in particular (it is by design the language of all the game's instructions, feedback, definitions, and tutorial messages), a knowledge of only 300 or 400 words in this zone could reasonably predict a tough time making sense of the game. Conversely, a score of over 800 may have given learners the impression they were already good at English (in the context of their school and community) and had little to learn. If this analysis is correct, it may also shed light on whether *Word Coach* can be used by both L2 and L1 populations. A first hypothesis is that for L2 learners, a score of 60% at the first level of the Size Test (600 of the first 1,000 words known) may be a precondition for play.

CONCLUSION

Does research-informed commercial electronic gaming have the potential to respond to the problem this study started with, that many learners are in serious need of a means of rapid vocabulary expansion? The results presented above suggest that *Word Coach* helped many of these learners develop their English lexicons on both declarative and procedural levels. The number of new words learned to meaning-recognition level amounts to an average vocabulary expansion of between roughly 10% and 20% in a four month period, while the base rate is probably something well under 5% per school year. And for the heaviest ten players in the present study (over 6,000 words fully recycled), the higher rates could translate into as many as 1,200 reasonably stable acquisitions in four months. Such an augmentation could make a big difference to an academic learner's lexicon, say from 1,800 to 3,000 known word families (or, from about 70% to 90% coverage in average academic texts). The learning power of electronic vocabulary games seems indisputable.

But it is increased access to old words that is in some ways the more interesting outcome of the study. Accessing forms from meanings, meanings from forms, and so on, under a time constraint, appears to be a language skill that can indeed be usefully practiced in a video game context. Interestingly, while none of the previous training studies in lexical access have yet shown any effect on language use (e.g., Fukkink, Hulstijn, & Simis, 2005), the significantly reduced code switching found in the present study may constitute such an effect.

On the down side, not all learners appeared to be entirely intrigued with *Word Coach*, despite the combined talents of language researchers and game designers that went into it.

And for those who persisted and made substantial gains, there is no proof they were not internally motivated, or at some point of learning readiness, such that they could have made similar gains with a well-designed set of word cards. It is not clear how many more learners would persist with an electronic game than would have persisted with similar learning principles embodied in a cheaper technology. The present study does not offer an answer to this question but it is probably worth asking.

Even more pressing are questions about how to best target which words should be learned, whether the needs of gaming and the needs of learning can be harmonized, and whether L1 and L2 learners can usefully inhabit a common platform. As mentioned, these questions come to a head on the matter of how the testing system calibrates the point of entry into the game for its two constituencies. It has not been shown that L1 and L2 cannot share a platform, just that the placement system needs to accommodate some L2 realities better.

Other questions raised and passed on to future game investigators include whether the hint of a 600-word entry point is correct, what duration of use is learning indicated, whether this period can be sustained by the game's motivational attractions, and what provisions for post-game follow-up can exploit and enhance gains. Such questions can only be answered with new studies that ask these questions specifically. Surely within *Word Coach's* three years of life and many thousands of players, many have now had a truly longitudinal experience of using the game, and if found and assembled they could furnish useful information.

A backdrop to this study was to investigate the trade-off between learning principles and access to a mass commercial playership. In this case, apart from the nonsensical designation of different lexical levels as "reporter," "lawyer," and "poet," etc., and some work to do for a common L1-L2 placement test, this particular company on the whole tempered the need to recoup its investment with an enthusiasm to get and use the best research information available in the design process.

All this would appear to augur well for the future of vocabulary gaming. Once some of the lingering questions have answers, the next step is probably to reinvest these findings into the development of a less structured, more contextualized type of video game with the sequencing and recycling of new words integrated into an immersive narrative format.

NOTES

¹ This invited symposium entitled *Getting the Word Out* was held at the American Association of Applied Linguistics conference in California in 2007 and was organized by Marlise Horst and Tom Cobb (see the complete PowerPoint presentation at http://www.lex tutor.ca/aaal_07/).

² The proposed baseline of 145 new words per year on the basis of two one-hour lessons per week (145 words/ (40 weeks x 2 hours) = 1.81 new words per hour) falls in line for plausibility purposes with Milton and Meara's (1995) finding that British school children learned 550 new words per year for five lessons per week (550 words/(40 weeks x 5 hours)=2.75 new words per hour) in a foreign language environment (British children learning French in a British school).

REFERENCES

- Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, 24(4), 425-438.
- Atkinson, R. C. (1972). Optimizing the learning of a second-language vocabulary. *Journal of Experimental Psychology*, 96(1), 124-129.

- Barcroft, J. (2007). When knowing grammar depends on knowing vocabulary: Native-speaker grammaticality judgements of sentences with real and unreal words. *The Canadian Modern Language Review*, 63(3), 313-343.
- Bates, E., & Goodman, J. (1997). On the inseparability of grammar & the lexicon. *Language & Cognitive Processes*, 12, 507-584.
- Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology*, 93(3), 498-520.
- Cambridge advanced learners' dictionary*, 2nd ed. (2005). New York: Cambridge University Press.
- Chall, J., & Jakobs, V. (2003). Poor children's fourth-grade slump. *American Educator*, 27(1), 14-15.
- Cobb, T. (1999). Breadth and depth of vocabulary acquisition with hands-on concordancing. *Computer Assisted Language Learning*, 12, 345-360.
- Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning and Technology*, 11(3), 38-63. Available at <http://ilt.msu.edu/vol11num3/cobb/default.html>
- Cobb, T. (2008). Necessary or nice? The role of computers in L2 reading. In Z. Han & N. Anderson (Eds.), *L2 Reading research and instruction: Crossing the boundaries* (pp. 144-173). Ann Arbor: University of Michigan Press.
- Cobb, T. (2010). Learning about language and learners from computer programs. *Reading in a Foreign Language*, 22(1), 181-200.
- Ebbinghaus, H. (1885/1913). *Memory: A contribution to experimental psychology*. New York: Teachers College, Columbia University.
- Elgort, I. (2007). The role of intentional, decontextualized learning in second language vocabulary acquisition: Evidence from primed lexical decision tasks with advanced bilinguals. Unpublished doctoral dissertation, Victoria University of Wellington, New Zealand.
- Ellis, N. (1994). Vocabulary acquisition: The implicit ins and outs of explicit cognitive mediation. In N. Ellis (Ed.), *Implicit & explicit learning of languages* (pp. 211-282). London: Academic Press.
- Fukkink, R., Hulstijn, J., & Simis, A. (2005). Does training in second-language word recognition skills affect reading comprehension? An experimental study. *The Modern Language Journal*, 89(1), 54-57.
- Gee, J. P. (2003). *What video games have to teach us about learning & literacy*. New York: Palgrave Macmillan
- Godwin-Jones, R. (2010). Emerging technologies from memory palaces to spacing algorithms: Approaches to second-language vocabulary learning. *Language Learning & Technology*, 14(2), 4-11. Available at <http://ilt.msu.edu/vol14num2/emerging.pdf>
- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.
- Gruneberg, M., Morris, P., & Sykes, R. (Eds.) (1978). *Practical aspects of memory: Current research and issues, Vol. 1: Memory in everyday life*. Chichester: John Wiley.
- Horst, M., Cobb, T. & Meara, P. (1997). Beyond A Clockwork Orange: Acquiring second-language vocabulary through reading. *Reading in a Foreign Language*, 11, 207-223.
- Hulstijn, J. H. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research*, 18(3), 193-223.
- Jiang, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal*, 88, 416-432.

- Laufer, B. (2005). Instructed second language vocabulary learning: The fault in the 'default hypothesis'. In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 286-303). Amsterdam: Mouton de Gruyter.
- Laufer, B. (2000). Task effect on instructed vocabulary learning: The hypothesis of 'involvement.' *Selected Papers from AILA '99 Tokyo* (pp. 47-62). Tokyo: Waseda University Press.
- Laufer, B., & P. Nation, P. (2001). Passive vocabulary size and speed of meaning recognition: Are they related? In S. Foster-Cohen & A. Nizgorodcew (Eds.), *EUROSLA Yearbook 1* (pp. 7-28). Amsterdam: Benjamins.
- Leech, G., Rayson, P., & Wilson, W. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. London: Longman.
- Mayer, M. (1967). *A boy, a dog, and a frog*. New York: Penguin Putman.
- Meara, P. (1980). Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching and Linguistics: Abstracts*, 13, 221-246.
- Meara, P. (1995). The importance of an early emphasis on L2 vocabulary. *The Language Teacher*, 19(2), 8-11. (See also http://www.lexutor.ca/research/meara_lists.mht)
- Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, 4, 142-54.
- Milton, J., & Meara, P. (1995). How periods abroad affect vocabulary growth in a foreign language. *ITL Review of Applied Linguistics*, 107/108, 17-34.
- Mondria, J.-A., & Mondria-De Vries, S. (1994). Efficiently memorizing words with the help of word cards and 'hand computer': Theory and applications. *System*, 22, 47-57.
- Mondria, J.-A., & Wit-de Boer, M. (1991). The effects of contextual richness on the guessability and the retention of words in a foreign language. *Applied Linguistics*, 12(3), 249-267.
- Nakata, T. (2011). A critical investigation of flashcard software for second language vocabulary learning. *Computer Assisted Language Learning*, 24(1), 17-38.
- Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59-82.
- Nation, P. (2001). *Learning vocabulary in another language*. New York: Cambridge University Press.
- Nation, P. (1990). *Teaching and learning vocabulary*. Rowley, MA: Newbury House.
- Nation, P., & Beglar, D. (2007). A vocabulary size test. *Language Teacher*, 31(7), 9-13.
- Oxford University Computing Services. (2000). *The British National Corpus*. Available at <http://www.natcorp.ox.ac.uk/>
- Pensky, M. (2006). *Don't bother me Mom - I'm learning*. St Paul, MN: Paragon House.
- Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second-language learning. In J. Kroll & A. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371-388). Oxford, UK: Oxford University Press.
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal*, 95(1), 26-43.
- Snellings, P., van Gelderen, A., & de Glopper, K. (2002). Lexical retrieval: An aspect of fluent L2 production that can be enhanced. *Language Learning*, 52(4), 723-754.
- Spector, J., & Ross, S. (2008). Introduction: Special thematic issue on game-based learning. *Education Technology Research & Development*, 56, 509-510.
- VanPatten, W. (1990). Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition*, 12, 287-301.

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review*, 57(4), 541-572.

AUTHORS' BIODATA

Tom Cobb pioneered language learning technologies in many ESL lands in the 1980s and 1990s. He now lives and works in Montreal, Canada, teaching courses and conducting research in technology development and applied linguistics. Tom's work comes together in the Compleat Lexical Tutor website (www.lextutor.ca), to which many language learners, teachers, and researchers worldwide use and contribute ideas. Lextutor, which is the basis of *Word Coach*, provides teachers with tools to support a data-driven and frequency based approach to language instruction.

Marlise Horst is an Associate Professor of applied linguistics in the Department of Education at Concordia University in Montreal, where she teaches courses in second language vocabulary acquisition, corpus linguistics, TESL methodology, and the history of English for language teachers. Her research interests include vocabulary acquisition through extensive reading and related computer assisted activities. Her recent research has investigated corpora of classroom speech and teacher talk. In the development phase of the *Word Coach* project, Marlise tested the game extensively, reaching the "expression potential" of a lawyer!

AUTHORS' ADDRESSES

Tom Cobb
Université du Québec à Montréal
Département de didactique des langues
C.P. 8888, Centre-ville
Montréal (Québec)
H3C 3P8 Canada
cobb.tom@uqam.ca

Marlise Horst
Concordia University
Department of Education
1455, De Maisonneuve Blvd. W.
Montréal (Québec)
H3G 1M8 Canada
marlise@education.concordia.ca