Electronic Dictionaries, Printed Dictionaries and No Dictionaries: the Effects on Vocabulary Knowledge and Reading Comprehension

by

Michael H. Flynn

A dissertation submitted to the
School of Humanities
of the University of Birmingham
in part fulfillment of the requirements
for the degree of

Master of Arts
in
Teaching English as a Foreign or Second Language (TEFL/TESOL)

This dissertation consists of 12,956 words
Supervisor: Stuart Webb

Centre for English Language Studies
Department of English
University of Birmingham
Edgbaston, Birmingham B152TT
United Kingdom

March 2007

Abstract

This paper describes an experiment with Japanese EFL university students comparing comprehension, and the receptive and productive vocabulary knowledge gained from reading an expository text with electronic dictionaries, printed bilingual dictionaries and no dictionaries. Vocabulary knowledge was tested with a two-week pre-test, an immediate post-reading test and a two-week post-test. Receptive knowledge was assessed with a checklist test for known and unknown items while productive knowledge was measured with a matching item test. Comprehension was measured with true/false questions translated into the students' first language. Dictionary usage tended to result in higher scores on the comprehension and vocabulary knowledge measures than students who read without dictionaries. Electronic dictionary usage tended to result in superior gains on the comprehension and vocabulary measures than printed dictionary usage. However, both groups of dictionary users required significantly more time to read the text than students who had not used dictionaries.

Acknowledgements

I would like to express my thanks to Stuart Webb for his guidance, suggestions and patience. I would also like to thank my students and Michael Raikes and his students for participating in test pilots and the experiment. I owe additional thanks to Terry Shortall and Nicholas Groom for reading and commenting on parts of the paper. I am most grateful to Simon Cole and Lachlan Jackson for initially suggesting that dictionaries would be an interesting area to research and to Peter Corr for his many insights into teaching. Finally, I would especially like to thank my wife, Keiko Fukuyama, who made this endeavor possible.

Contents

Tables

1 Introduction

This study has its origins in educator, teacher, and researcher views of dictionary use by second and foreign language learners that range from ambivalence to outright antipathy. Knight, however, argued that many reservations about dictionaries and the resulting teaching practices were not backed by empirical evidence (1994: 285-6). She conducted an experiment that compared incidental receptive and productive vocabulary learning and reading comprehension of students reading short texts on a computer screen with and without access to dictionary definitions accessible through the computer text. Nonetheless, dictionary studies remain scarce or limited in scope and Knight's findings have yet to be replicated with other types of dictionaries, language learners, and texts with different characteristics. The present study is loosely modeled on Knight's study where the basic research issues are whether students gain more words from guessing from context or dictionaries and how dictionary use affects reading comprehension.

After presenting relevant research in commonly held views of dictionary use, related dictionary studies, vocabulary learning, vocabulary knowledge and reading comprehension, this paper describes the design, and reports and discusses the results of an experiment with Japanese EFL learners which compares performance on receptive vocabulary, productive vocabulary and comprehension measures of a group reading a text without dictionaries to two groups reading the text with access to the two types of dictionaries likely to be found in EFL classrooms--the traditional printed bilingual dictionary (PD) or the hand-held electronic dictionary (ED).

2 Related Research on Dictionaries, Vocabulary and Reading

2.1 Views of Dictionary Use

Nist and Olejnik (1995: 172) ask the question, "where has this idea come from that looking up words in the dictionary is the worst way for students to learn vocabulary?" Some EFL teachers discourage use of both monolingual and bilingual dictionaries in the belief that dictionaries do not help students to understand vocabulary in context and because students overuse dictionaries at the expense of developing the ability to guess from context and self-confidence (Bensoussan, Sim and Weiss 1984: 262) while others advocate only using the target language and are concerned that bilingual dictionaries used for word for word translations will adversely affect student comprehension at the sentence and discourse level (Tang 1997:39). According to Snell-Hornby (1984) and Yorkey (1970) reported in Aust, Kelley & Roby (1993: 66), "...many language educators... believe that bilingual dictionaries are counter productive because they cultivate the erroneous assumption that there is a one-to-one correspondence between the words of the two languages." Because monolingual dictionaries may be seen as solving some of the problems presented by bilingual dictionaries, most teachers prefer the monolingual dictionary (Koren, 1997:2). However, it may be difficult for a student with insufficient vocabulary to understand a monolingual dictionary entry that contains unknown words and time-consuming or even frustrating if understanding the entry requires looking up other entries with still more unknown words. Learners can also misinterpret monolingual dictionary entries and the entries can be misleading (Nesi and Maera, 1994 in Koren, 1997: 2-3). Modern electronic pocket dictionaries (ED) can enable students to look up words 23% faster than conventional dictionaries (Weschler and Pitts, 2000: 1), but the increased speed of ED lookup may be at the expense of engagement and deeper processing of the words possibly resulting ultimately in less vocabulary learning (Stirling, 2003: 2-3). Stirling (ibid.) also conducted a small survey of EFL teachers who listed, "insufficient examples, inaccurate meanings, unintelligible pronunciation, lack of collocations, excess of meanings, and the absence of improvements found in other dictionaries" as possible disadvantages of ED. Knight (1994: 285) includes another concern of educators that may apply to all types of dictionaries,

"looking up words frequently interferes with short term memory and thus disrupts the comprehension process".

2.2 Studies of Dictionaries and Reading Comprehension

Studies have not been able to establish that using a dictionary consistently improves reading comprehension. Bensoussan, Sim and Weiss (1984) examined the effect of bilingual and monolingual dictionaries and no dictionary on reading comprehension of Israeli EFL university students with multiple choice questions in a variety of text passages. No significant differences were found in reading comprehension or time required between the control groups and the dictionary groups. Most students did not look up very many words. One conclusion was that, "less proficient students lack the language skills to benefit from a dictionary, whereas more proficient students know enough to do without it (ibid.: 271).

Koyama and Takeuchi (2004b) compared handheld electronic dictionaries (ED) and printed bilingual dictionaries (PD) on reading tasks with 72 Japanese EFL university students. Two texts were read while using dictionaries. PD users spent 16% more time reading than the ED group but depending on which text was read, the ED users looked up anywhere from 5.5 times to 1.7 times as many words. Both results were statistically significant. However, on six-question multiple choice comprehension test scores, there was no statistically significant difference. Unfortunately, the study does not provide the texts or the quiz questions used and there was no comparison to a group not using dictionaries. The types and the small number of questions used may not have adequately measured differences in comprehension.

Albus, Thurlow, Liu, and Bielinski (2005) used a simplified monolingual English dictionary and compared the effects on comprehension of a newspaper article for Hmong ESL learners and native speaker Junior High Students. Overall, they did not find any significant difference in scores between no dictionary control and dictionary groups but reported that 59% of students in the ESL group did not use their dictionary or used it only for a few words. Of the high, intermediate, and low level students in the dictionary group

that did report using dictionaries, only the intermediate group showed a significant score difference. The results of this study are similar Bensoussan, Sim, and Weiss (1984), in that many students did not extensively use their dictionaries and because high and low level students did not benefit from the dictionary. The result with the intermediate students hints that for comprehension scores to be affected by dictionary use that there must be an intersection between reader ability and the text such that the text is neither too easy nor too difficult.

Knight (1994) conducted an experiment that compared incidental receptive and productive vocabulary learning and reading comprehension of second year Spanish as a foreign language U.S university students reading 250 word authentic texts with 95.2% known words on a computer screen with and without access to dictionary definitions accessible through the computer text. After reading texts, students wrote a recall summary to check comprehension. Compared to the no dictionary group, students using dictionaries attained significantly higher scores. Comprehension scores were further analyzed by dividing students according to high and low ability. Both ability groups had higher scores in the dictionary condition but only the low ability group showed a statistically significant increase over the no dictionary group. Additionally, correlations between number of look ups and reading comprehension varied according to student ability with low ability students having a high correlation between comprehension and the number of look ups while high ability students had a much lower correlation. In other words, looking up more words helped the comprehension of low ability students more than high ability students. Finally, the dictionary group was found to require roughly 42% more time to read than the no dictionary group. Compared to the no dictionary group, the high ability group's scores were 18% higher while the low ability group's scores were 45% higher indicating that only the low ability group had an increase in comprehension commensurate with the additional time incurred by using a dictionary. The Knight study suggests that the intuitive notion that dictionary use will lead to improved comprehension only holds true under certain conditions.

Research on dictionary use and comprehension suggests a number of difficulties in observing improvements in comprehension resulting from dictionary use. Some tests used to measure comprehension may simply be inadequate for the task while the texts may be too easy or too difficult for the ability level of the readers for dictionary use to make a difference in comprehension. Because of the difficulty of objectively scoring large numbers of written recall summaries used in Knight's study and the unavailability of a second rater to establish reliability, in the present study an attempt has been made to develop a simpler comprehension measure that will be sensitive enough to capture comprehension differences in a difficult authentic expository text containing less than 95% known words. As discussed in section 2.4, the risk of going below the 95% level is that even with dictionary use, the text will simply be too difficult. It is possible however that by using a text with a lower number of known words that differences between context only and dictionary use will become more apparent if dictionaries can be used to bootstrap up to a higher percentage of known vocabulary sufficient to permit comprehension.

2.3 Studies of Dictionaries and Vocabulary Learning

There is a tendency for dictionary studies to show that dictionary use leads to vocabulary gains. Luppescu and Day (1993) studied 293 Japanese EFL university students using no dictionaries or printed bilingual dictionaries and compared vocabulary acquisition and time taken to read a five page narrative edited to contain enhanced content and multiple occurrences of target words to assist students in guessing. The dictionary group took twice as long to read but only acquired a 50% greater mean score on a multiple choice vocabulary quiz. However, for some items with multiple dictionary definitions, the no dictionary group performed better than the students using dictionaries.

In Knight (1994) students read texts for meaning and wrote summaries without being informed about immediate post-reading vocabulary tests. The first vocabulary task was supplying equivalent L1 words or definitions for target words. The second task required selecting L2 definitions from multiple choice items for each L2 target word. The vocabulary tests were subsequently administered again two weeks latter unannounced.

Although the direction of the first task is from L2 to L1, Knight considers this task to be productive. However, because of the L2 to L1 direction of the task, it might more appropriately be considered a test of receptive knowledge with a recall component (See Nation, 2001:29-30). Analysis was conducted for no exposure to text or dictionary, exposure to text without the dictionary, and exposure to the text with access to the computer dictionary. Students using the computer dictionary attained statistically significant higher scores on both the immediate and delayed vocabulary measures of both vocabulary tasks. Furthermore, high ability students learned more words than low ability students. Low correlations were found between number of look ups and the supply definition scores but the number of words looked up correlated highly with the select definition scores. Looking up a larger number of words does not appear to have interfered with vocabulary acquisition and helped in the case of the supply definition questions. Finally, the percentage gains of the low ability students on the immediate and delayed vocabulary tests for both vocabulary tasks exceeded the learning attributable to the 42% increased time on task from dictionary use. In contrast, the percentage gains of high ability students were less than or roughly equal to the percentage increase in time on all measures except the immediate supply task. Thus, low ability students benefited more from the time spent looking up words.

Nist and Olejnik (1995) studied 186 U.S university students who were given 20 minutes to study 10 artificial words presented in short contexts of a couple sentences followed by dictionary definitions. The quality of the contexts and definitions was manipulated to create strong and weak conditions. Participants were tested on receptive and productive vocabulary measures. The primary finding was that students performed significantly better when they were exposed to strong definitions regardless of whether they were exposed to the word in strong or weak contexts. This suggests that more will be learned from a dictionary with good definitions than from context alone.

In Koyama and Takeuchi (2004a), 18 Japanese EFL university students read an English text without dictionaries and used ED or PD to look up and write four word definitions and four usage examples for eight target words. There were no significant differences in the search

times between the ED and PD conditions. The responses were scored to determine how well students had used the dictionaries to obtain information and then compared with recall and recognition tasks administered seven days latter. Students using the PD achieved higher average scores on both recognition and recall. The difference on the recognition scores was statistically significant and was attributed to greater depth of processing required by the look up procedures in the PD,

> EFL learners were obliged to do an arduous or elaborate work in the process of searching in the PD condition, while they could easily get a word definition only by inputting the spelling of the word in ED condition. (ibid.: 42)

One question in regard to this study is whether students would experience differences in recall and recognition if the look ups had been performed during the reading in response to the students' self-directed attempts to comprehend the text.

In summary, the studies presented here appear to indicate that compared with comprehension, it is easier to observe gains in vocabulary as the result of dictionary usage. There is also evidence that dictionary use may lead to more words learned than from context alone. While the present study is similar to Knight's study, a simpler methodology requiring only one text and less complex statistical calculations was employed. Additionally, different types of tests for vocabulary gains were used in an attempt to improve measurement sensitivity.

2.4 Minimum Vocabulary for Contextual Learning

In any experiment involving learning from context both the student's vocabulary and the vocabulary of the text should be taken into account. Read (2000: 74-5) states that "[there is a] well-documented association between good vocabulary knowledge and the ability to read well". Learning words while reading requires a certain minimum vocabulary knowledge. Laufer and Sim (1985) in Nation state that 65-70% of vocabulary must be known to comprehend texts for academic purposes and that, "...the most pressing need of the foreign language learner was vocabulary, then subject matter knowledge, and then syntactic

structure" (2001:145). Johns (1980) in Bensoussan, Sim and Weiss suggest that if 5% or more of the words are unknown that comprehension of the text structure and guessing from context may not be possible (1984: 264). Hu Hsueh-chao and Nation (2000) compared the effects on comprehension of different levels of known words in fiction texts. Their results, summarized in table 2.1, show that even small amounts of unknown vocabulary adversely impact comprehension. Section 3.4.1 describes three methods used to estimate the amount of unknown vocabulary for the text used in this study.

Table 2.1 **Words Known Versus Adequate Comprehension**

Percentage of known words in text	Percentage of Subjects achieving adequate comprehension (12 out of 14 questions correct on a multiple- choice measure)
80%	0%
90%	25%
95%	35%
(98%) No text with 98% coverage was used. This number was derived by interpolation using a jittered plot graph.	Most learners gain adequate comprehension.
100%	88%

2.5 Guessing from Context

Even when a text does contain a sufficient percentage of known words for general comprehension, students unaided with other resources need to guess the meanings of new words. Bensoussan and Laufer (1984) studied lexical guessing in context with a 574 word text containing 70 target words. Of these, 29 did not have any contextual clues and of the 41 remaining only 13 were clearly defined in the text while the other 28 words had indirect clues such as collocations, contrasts, or word pairs (ibid. 21). In many instances, words may not have sufficient context to permit successful guessing. Lexical guessing was only successful in "13% of responses for 24 percent of the total words" (ibid.: 25). The most frequent guessing errors were incorrect choice of a word with multiple meanings (20%), mistranslation of a morphological trouble maker (17%), Mistranslation of an idiom (16%), confusion with similar sounding L2 words (13%), confusion with a similar sounding L1

words, and wild guessing (11%) (ibid.: 23). Furthermore, student ability did not affect the success rate of guessing,

> Weak, average and good students all applied the same strategies: ignoring words, using 'preconceived [and incorrect] notions'…although better students may *know* more words, they do not guess differently or guess more than weaker ones. (ibid.: 26)

To measure the learning of words from context from a single exposure, Nagy, Herman and Anderson (1985) studied native English speaker children reading two 1,000 word texts, one a narrative and the other expository. Depending on the test format and difficulty level of the test, the probability of learning a word ranged from approximately 10 to 20%. In a similar study, Nagy, Anderson, and Herman administered tests six days after the reading passage where the probability of learning a word from context was only 5% (1987: 261). Given the low success rates in guessing, and the presence of words that are impossible or difficult to guess from context, many texts, even when suited to the ability of the students, will still contain words that can only be learned from extra-textual references such as dictionaries, glosses, or teacher explanations.

2.6 Incidental and Intentional Vocabulary Acquisition

In spite of the difficulties of guessing from context, people do manage to learn vocabulary in both their native and foreign languages. The question that arises at this point, then, is how does this process take place? One view is that learning can be divided into incidental learning and intentional learning. Nation defines learning from context as:

> …the incidental learning of vocabulary from reading or listening to normal language use while the main focus of the learners' attention is on the message of the text. Learning from context thus includes learning from extensive reading, learning from taking part in conversations, and learning from listening. Learning from context does not include deliberately learning words and their definitions or translations even if these words are presented in isolated sentence contexts. (Nation, 2001: 232-33)

Arguably this is consistent with Krashen who equates incidental learning with his Input Hypothesis where language is subconsciously acquired while the conscious focus is on message and intentional learning with his Monitor Hypothesis where conscious focus is on

form. Of the two, he argues that there are, “severe limits on how much can be learned [from intentional learning]” (1989: 440) and that vocabulary size of native speakers and the mastery of the complex properties of the vocabulary are too great to be accounted for by conscious learning (ibid.: 452-3). Krashen recommends that vocabulary learning should take place through, “massive quantities of pleasure reading” (ibid: 455).

Nagy concurs that for native speakers the bulk of vocabulary acquisition results from incidental learning during extensive reading and that instruction plays a much smaller role. He argues that while the chances of acquiring a new word through context in a single encounter are small, that small gains combined with large amounts of reading result in large numbers of new words being learned. He calculates that if an average student reads a million words a year with 2% of the words being unknown (20,000 unknown words), that a 5% chance of learning a word would yield an annual gain of 1,000 words per year (1997: 75). The ‘book flood’ studies examined by Elley (1991) also suggest that young second language learners benefit from extensive reading when the focus is on meaning and not on form. However, Hill and Laufer (2003) argue that reading millions of words is not an entirely plausible explanation or solution to the problem of learning the first couple thousand words of a second language,

> This would appear to be a daunting and time consuming means of vocabulary development. It seems therefore reasonable that L2 learners acquire their vocabulary not only from input, be it reading or listening, but also through word focused activities. (Hill and Laufer, 2003: 88)

In the case of Japanese EFL students, my own experience suggests that both word focused instruction and intensive reading of difficult texts in preparation for university entrance exams are more likely to account for a substantial percentage of basic vocabulary knowledge than extensive reading.

One problem with the notion of incidental vocabulary learning is that it may simply be a researcher construct that does not reflect what actually happens. For example, Paribakht and Wesche (1999) used introspective think aloud protocols to examine learning words from

context and found that subjects relied on sentence level grammar, word morphology, punctuation, world knowledge, homonymy, word associations and cognates and that learners guessing strategies were influenced by the type of comprehension tasks, readers' perceptions of the text difficulty and interest, individual differences in learner strategy use, and views on the value of reading for vocabulary learning. They state that, "vocabulary learning through reading is in some sense not "incidental," at least from the learner's perspective" (ibid.: 215). Gass goes further and argues that it is not possible to know whether a word has been learned incidentally (1999: 319). Even when a reader is focused on meaning, considerable conscious attention and cognitive effort may be directed to understanding new words.

In the present study, students are told that they will be tested for comprehension without being told that they will also be tested on vocabulary. Even so, the experimental setting may invalidate the argument that learning is incidental. Additionally, using dictionaries clearly involves conscious cognitive resources that would probably best be classed as intentional learning. Furthermore, the reading task in the present survey contains a considerable number of words not known to the readers and is probably best considered as intensive reading which is not the type of pleasure text recommended for incidental learning in extensive reading programs (Hu Hsueh-chao and Nation, 2000: 423). For these reasons, it is important to recognize that the "incidental" learning taking place in this study should primarily be considered incidental in reference to the task directions that the students read for meaning.

2.7 Vocabulary Knowledge

Many vocabulary tests give the impression that a correct response indicates that an item is "known". However, the actual state of knowledge for any given item is more complex. Henriksen describes vocabulary knowledge as occurring in three dimensions: partial-precise, depth, and receptive-productive (1999: 304). Precise knowledge is exemplified by tests, "requiring the ability to translate the lexical item into the L1, to find the right definition in a multiple-choice task, or to paraphrase in the target language" (ibid.: 305). In contrast,

checklist tests that simply require a word to be checked if it is "known" attempt to include measures of partial knowledge such as word recognition (ibid.: 305). Depth of knowledge refers to knowing multiple meanings and senses of a word, the relationships with other words, knowledge of collocational features, and factors related to when a word is used (ibid.: 305-306). The final receptive-productive dimension refers to, "… ability to use the words in comprehension and production" (ibid.: 307).

Richards (1976: 83) in Read (2000: 25) identified various knowledge types necessary to fully acquire a word:

> Knowing a word means knowing probability of encountering a word in speech or print...the sorts of words most likely to be associated with the word...the limitations on use of the word according to variations of function and situation...the syntactic behavior...underlying form and derivations...the network of associations between that word and other words...the semantic value of a word...the different meanings associated with the word. (Richards, 1976: 83)

It should be evident that depth of knowledge is multi-dimensional, might include other types of knowledge such as etymology and is unlikely to be measurable by a single test item. Related to the notion of depth is the idea that learners may pass through stages of knowledge on the way to acquisition. Wesche and Paribakht (1996) in Maera (1999) use a set of tests with a scale of five stages progressing from no knowledge of the word up to productive ability:

> 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. eg: _____. (Wesche and Paribakht, 1996)

Maera argues that the stages should not be viewed as progressive since, for instance, it is possible to produce a sentence without knowing what a word means and that knowledge at the different stages is unstable and may change over time. He also notes that although there are other possible stages and no account for acquiring multiple meanings, scoring of even the five-item test is likely to require considerable time and effort and that the number of

vocabulary items that can be tested will be limited (1999:6). In the present study, the focus is on measuring the learning from single encounters in one text. Under these circumstances, it is assumed that depth of knowledge will be limited.

The receptive-productive (R/P) distinction according to Melka is an intuitive construct based on the assumptions that productive vocabulary is smaller than receptive vocabulary, that productive ability is acquired after receptive ability and "the fact that language users (especially children), understand novel derived forms before they can produce them" (1997: 84). However, she states that, "Knowing a word is not an all or nothing proposition; some aspects may have become productive, while others remain at the receptive level" (ibid.: 87). Nation, (2001: 27) illustrates this in Table 2.2 by applying the receptive/productive distinctions to word knowledge of form, meaning, and use.

Table 2.2 **Categories of Receptive and Productive Knowledge**

Form	Spoken	R	What does the word sound like?
		P	How is the word pronounced?
	Written	**R**	**What does the word look like?**
		P	How is the word spelled?
	Word parts	R	What parts are recognizable in this word?
		P	What word parts are needed to express this meaning?
Meaning	**Form & Meaning**	**R**	**What meaning does this word signal?**
		P	**What word form can be used to express this meaning?**
	Concept & Referents	R	What is included in the concept?
		P	What items can the concept refer to?
	Associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	Grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	Constraints on use (register, frequency)	R	Where, when, and how often would we expect to meet this word?
		P	Where, when, and how often can we use this word?

The aspects of R/P knowledge tested in the present study are in bold in the table. While Nation's form & meaning category only represents a small portion of potential R/P

knowledge and Nation views full learning of a word as an incremental process requiring multiple encounters, the present study attempts the difficult task of measuring knowledge gained from single encounters with a word while using inferencing strategies or dictionaries. Consequently, this study focuses on capturing learning of a basic R/P meaning rather than more complex types of knowledge. Nagy, Herman, and Anderson (1985: 237) provide some encouragement in noting that, "Although a single encounter with a word would seldom lead to full knowledge of its meaning, we believe that a substantial, if incomplete, knowledge about a word *can* be gained on the basis of even a single encounter."

2.8 Reading Comprehension

Vocabulary knowledge may be the most important single factor in reading comprehension. According to Alderson (2000: 99), "In studies of readability, most indices of vocabulary difficulty account for about 80% of the predicted variance". However, vocabulary knowledge while important is not the only factor affecting comprehension. Alderson (ibid.: 32-84) surveys some of the factors that affect reading and divides them into the reader and text variables summarized in Table 2.3.

Table 2.3 **Reader and Text Variables**

Reader Variables	Text Variables
Knowledge: lexical, syntactic, rhetorical, metalinguistic, discourse, L1 vs L2 knowledge, genre/text type, subject matter/topic, world, cultural Motivation: intrinsic/extrinsic Strategies Skills Purpose: scanning, skimming, rauding, learning, memorizing Real World vs. Test Taking Affect Stable characteristics: sex, age, personality Physical characteristics: eye movement, speed of recognition, automaticity of processing	Topic and Content Genre Organization Linguistic variables Readability Typographical features Verbal vs. Nonverbal information Medium of text presentation

These variables interact with each other in the process of reading but because there are so many variables, it is difficult to do justice to a discussion of reading comprehension in the

space permitted. Nonetheless some of the reader and text variables specific to this study are described in the section 3.1 on participants and section 3.4 on materials.

2.9 Research Questions

The design, materials, and dependent measures of this experiment take into account a wide variety of considerations which incorporate research in the minimum vocabulary necessary for contextual learning, guessing from context, the difference between incidental and intentional learning, the different types of vocabulary knowledge, and the variables involved in reading comprehension. The experimental research questions specific to this study are:

1. Are there significant differences in pre and post-reading receptive and productive vocabulary measures between the exposure to text only, exposure to text and printed dictionaries (PD), exposure to text and electronic dictionaries (ED), and no exposure to text or dictionary groups?
2. Are there significant differences in reading comprehension scores between the exposure to text only, exposure to text and PD, exposure to text and ED and no exposure to text or dictionary groups?
3. Are there significant differences in the time taken to read the text between the exposure to text only, exposure to text and PD, and exposure to text and ED groups?

3 Methodology of an Experiment to Measure Vocabulary and Comprehension Performance of Students Reading a Text with and without Printed or Electronic Dictionaries

3.1 Participants

The participants were 174 first and second year Japanese EFL students who had studied English in Junior High and High School. The students are assigned to classes by TOEFL scores ranging from 577 to 350 but these scores were not available for individuals. To obtain an individual ability measure, the Nation 3000 Word Level Vocabulary tests were administered (See Nation 2001: 418-19). A maximum of 30 points was possible. The average of all participants was 21.54. The Standard Deviation was 3.97.

3.2 Design

First and second year Japanese EFL university students from eight classes were given a receptive vocabulary checklist test (RVC) and a productive vocabulary definition selection test (PVDS) two weeks prior to the experiment. On the day of the experiment, students were divided into a group that read an authentic expository text without dictionaries and two groups that read the text using either PD or ED. A fourth group did not read the text or use dictionaries but took the RVC and PVDS to investigate whether scores would increase over time due to students looking up words after tests on their own initiative, noticing target words and learning them between tests, or even from simply repeatedly taking the tests. Immediately after the reading, the RVC and PVDS were administered a second time and followed with a comprehension test. The no text no dictionary group (NTND) also took the comprehension test to determine the effects of background knowledge on the comprehension test. Two weeks after the experiment, the RVC and PVDS were administered a third time to all groups. Table 3.1 shows the order of the tests administered to each different group.

Table 3.1 **Groups and Order of Test Administrations**

Groups	Pre-experiment test (2 weeks prior	Immediate Post-experiment test	Delayed Test 2weeks after experiment
No Text, No Dictionary **(NTND)**	RVC->PVDS	RVC->PVDS ->Comprehension	RVC->PVDS
Text, No Dictionary **(ND)**	RVC->PVDS	RVC->PVDS ->Comprehension	RVC->PVDS
Text plus Electronic Dictionary **(ED)**	RVC->PVDS	RVC->PVDS ->Comprehension	RVC->PVDS
Text plus Printed Dictionary **(PD)**	RVC->PVDS	RVC->PVDS ->Comprehension	RVC->PVDS

3.3 Procedures

During the RVC and PVDS two-week pre-tests, participants were told that they were receiving a level-check. They were not informed that they would be taking the same tests again. For the experiment, students in the ED, PD and ND groups were instructed to read the text for meaning and told they would take a comprehension test after reading but would not have access to the text after starting the test. Participants were not informed that they would also be tested on vocabulary in the RVC and PVDS. When students finished reading, they raised their hands so that the reading time could be recorded and to receive the first test. The ED and PD groups were permitted to use dictionaries during the reading but not while taking the tests. Groups using dictionaries underlined words that they looked up. Students not receiving a text took the same tests as the other groups but were told that that they were being tested on guessing ability and background knowledge. Tests were collected as they were finished and new tests distributed. Participants were not informed about the two-week post-test.

When students finished any test series, they were given individual learning tasks unrelated to the experiment where dictionaries were not allowed to discourage immediate post-test look up of items in the experiment and so that early finishers would not be disruptive. None of the target words appeared in the tasks.

The order of the tests attempts to minimize learning from the tests. So that students could not change answers or refer to earlier test forms, students only had possession of one test at a time—the RVC, the PVDS or the Comprehension test. The RVC contains only the target words without definitions. The PVDS contains definitions but learning vocabulary from the test should be difficult because of the large number of distractors. On the day that the texts were read, the comprehension test was given last because it contains L1 translations that are in many cases quite similar to passages of the L2 text and might provide clues about unknown words.

Approximately one week prior to the experiment, students were asked to bring both a PD and ED if they owned both or either if they only owned one type of dictionary. Assignment into groups was not completely random. Whether the students had brought a dictionary or what type of dictionaries they brought was taken into account to create groups of roughly equal size. However, because of the concern that absences might jeopardize the experiment if there were an insufficient number of participants in the post-tests in the ND, PD and ED groups, a smaller number of students was assigned to the NTND group. Students were only allowed to use their own dictionaries. Because the dictionaries are owner-supplied in this study, the dictionaries in each group are not homogeneous which represents a confounding variable. However, it is believed that the dictionaries of each group will have more in common with each other than the dictionaries in the other group.

3.4 Materials

During the experiment, students read the text in Appendix I which is an excerpt from the book, The McDonaldization of Society, Ritzer, (1996: 106-7). Selecting an appropriate text was particularly difficult and involved a considerable degree of intuition. Ultimately however, any experiment on reading will not be able to escape a variety of effects resulting from the particular text(s) used which may affect the degree to which experimental results can be generalized. Nonetheless, all students involved in the experiment read the same text and were subject to the same text variables described in section 2.8. Table 3.2 describes the McDonaldization text in terms of these variables.

Table 3.2 **Text Variables Specific to Text**

Text Variable	**Characteristics of Excerpt from The McDonaldization of Society**
Topic and Content	Education and Sociology: negative effects of fast food style management
Genre	Expository argumentation, borderline between academic and popular writing.
Organization	Six paragraphs which usually include a general statement followed by examples and conclusion
Linguistic Variables	Text contains approximately 90% known vocabulary.
Readability	663 words. 5.4 characters/word
Typographical features	12 Point, single spaced, Century font
Verbal vs. Nonverbal Information	Text only, No pictures or illustrations.
Medium	Printed on A4 sheets of paper

3.4.1 Unknown words in Text: Three Approaches

Another important consideration was how much of the text's vocabulary was known to the subjects. As an initial pilot step towards investigating this, three students were given a copy of the text and asked to scan the text and circle unknown words. This required ten to fifteen minutes. The results are shown in Table 3.3.

Table 3.3 **Percentage of Unknown words in Pilot of 3 Students Scanning Text**

	Student X	Student Y	Student Z
Unknown words	45	30	32
Percentage known	93%	95%	95%
Average	95%	95%	95%

These percentages are roughly equal but probably underestimate the quantity of unknown words. Two of the students probably did not consider the title to be part of the article and may have not included the word "Docile". Also, it is impossible to estimate to what extent the students under claimed unknown words on the full text versions. Another interesting point was all three students identified "tyranny" as an unknown word. However, one student also underlined, "tyranny of the clock". Only "tyranny" was counted as an unknown word but at least one student viewed the words as an unknown chunk. Even though students were instructed just to skim and underline unknown words, all three students probably also read for meaning in varying degrees. Some words that were previously unknown may have been inferred from context and not selected as unknown (Read 2000: 156). Other words that

may have seemed less significant to understanding the meaning may have also not been underlined as unknown. Finally, the sample of three is quite small.

To provide another method of estimating unknown vocabulary, a reduced checklist of lexical items in each article was developed by deleting extremely frequent words like a, the, be, do etc., easy words that would be known by Japanese High School students, and words widely borrowed into Japanese. When a singular and plural word both appeared in an article, plurals were eliminated unless the only example was already in the plural. Third person verb forms were eliminated unless the only example was in the 3rd person. To determine whether students under claimed unknown words, ten imaginary words that resemble English words were introduced into each instrument. Meara and Buxton (1987) use a similar procedure, where students circle words that they claim to know. To avoid the words making a coherent story in spite of deletions of intervening words, the word list was partially scrambled. The finalized checklist was administered to a sample of ten students. The number of students that indicated that a particular word was unknown is tallied in Table 3.4. Imaginary words are shown in bold. The number following a word is the number out of ten students that indicated that the word was unknown.

Table 3.4 **Tally of Unknown Words from Pilot Checklist (Ten Students)**

---

spontaneity10 education0 docile8 developed0 variety0 nonhuman1 exert6 **snickle**6 process0 “for instance”3 follow0 hire1 enormous7 determine0 **minurite**10 leave0 educational0 assigned2 “no matter”1 lecture1 requires0 **servate**8 grading1 submitted4 employ0 “computer-graded”3 “multiple-choice”9 required0 **undiration**10 evaluations5 force0 lead0 ratings5 publishing0 regulations3 “time-consuming”8 **recund**10 tenure9 devote2 “of course”0 controlled0 system0 “for example”0 expert1 certain0 happens0 promotion1 **glinder**10 constraints7 mentioned1 leeway10 “highly structured”5 perform0 specific2 customers0 actually1 “grade schools”2 “in particular”4 strive8 described1 “boot camp”7 labeled4 mechanisms4 authority6 embrace6 **blossly**8 rationalized10 procedures9 rote5 objective2 creativity0 tend1 rewarded6 conform3 fascination5 discouraged3 leading0 docility10 thus5 tyranny10 focus2 find0 cluster7 excited0 examining1 **dinated**9 intensity9 turtle1 insists1 science0 employees1 crabs4 overall4 emphasis5 submissive8 kindergarten5 malleable10 creative0 independent0 rules0 “point of view”1 messy5 extreme3 appears0 demands3 equivalent10 “short-term”1 training0 “child care”1 largely0 **julique**10 exam0 obey4 determined1 instruction1 care0 curriculum3 activities0 “spelled out”5 detail1 clearly0 skilled0 experienced0 “McChild”6 seek0 comprehend5 producing0 plan0 rather0 “end up in”6 version0 abrivator10 relatively1 untrained2 technology0 omnipresent10 franchised6 MacDonald’s0 centers0 remedial9 corporation1 trains1 tailors7 uniformity3 “U-shaped”6 charges1 methods1 “for-profit”6 Sylvan5 “ready-made”6

The tally provided an indication of which words should be focused on in the vocabulary tests. However, the tally cannot be directly used to determine the percentage of unknown words in the text. The tally counts Types not Tokens. The problem of how to count unknown words is further complicated by multi-word items. None of the three students that received the full text version underlined any of the words in the phrasal verb “end up in” as unknown but on the decontextualized unknown word test, 60% indicated that they did not know the meaning of the phrase. Furthermore, if students had been presented with the words “end”, “up” and “in” separately they probably would have indicated that none of the words were unknown. These issues aside, each unknown phrase was counted as a single word.

To use the checklist tally to obtain a token count, the number of occurrences of each word in the text was tallied separately. The number of students that chose an item as unknown was then multiplied by the number of occurrences in the text and the results totaled. The word Sylvan was excluded because it is a proper noun. This method yielded 455 unknown words between ten students or an average of 45.5 unknown words per student or approximately 6.9% unknown words in the text. However, this still doesn’t contain a correction for students that under claimed unknown words. To get an approximate correction, the number of identified unknown words 455 was multiplied by the total number of imaginary words in the test (100) and the total was divided by the number of imaginary words that students that the students correctly identified as unknown (94) for an adjusted number of 484 words or approximately 7.3% unknown words in the text. The difficulty with this adjustment is it isn’t possible to know precisely which words students didn’t claim or how many times the unclaimed unknowns occurred in the text.

To avoid a correction for unclaimed unknown words, students that did not correctly claim all ten imaginary words as unknown were excluded. The average number of unknown types indicated by the seven students that correctly claimed all ten as unknown is 44 which is close to the average of 45.5 for all ten students. Each real unknown item was multiplied by the number of occurrences in the text. The result is displayed in Table 3.5.

Table 3.5 **Number and Percentage of Unknown Tokens**

Participant number	1	2	3	4	5	6	7
Number of unknown real tokens	55	52	40	44	78	51	43
Percentage of unknown words	8.3	7.8	6.0	6.6	11.8	7.7	6.5

Average number of unknown Tokens: 52. Average % of unknown tokens in text: 7.8%

Each of the methods has disadvantages. Recognizing words out of context is more difficult than recognizing them in context. However, providing context gives the reader an opportunity to infer words, which alters the number of unknown words reported. Additionally, just as there are different knowledge levels and types of knowledge about words, some students may have differing interpretations of what it means to know a word. A couple of students in particular were responsible for most of the words that were selected only once as unknown. Rather than having a major difference in vocabulary as compared to the other students, these students may have had a stricter internal definition of what it means to know a word than the other students. It is also possible that some of the words excluded from the decontextualized word list might have been claimed as unknown. However, based on the similarity of the results from the three different methods, the text is estimated to have roughly between 90 to 95% unknown words. Further taking into account that multi-word units were counted as one word would likely put the estimate closer to 90% if lexical items like “boot camp” were counted as two rather than one unknown word.

3.4.2 Location and Importance of Unknown Words in Text

The final step in selecting the text was investigating the location of unknown words. The two main concerns were that unknown words might be clustered in one section or only occur in areas not central to understanding the text. The writing is essentially expository following a pattern of general statements supported by details, arguments and statements that draw conclusions. Sentences containing general statements, main arguments, conclusions and the title were categorized as main idea sentences while the remainder were categorized as detail sentences. For ease of visual processing, a combination of colors, text

sizes, and italics was used to code the text (See Appendix II) for the locations of unknown words and whether the words occurred in main idea or detail areas.

The information in the text is further summarized in Table 3.6 which indicates that the words chosen by 7-10 students as unknown are distributed relatively equally throughout the text. Additionally, 45% of the words occur in main idea sentences with the remainder in detail sentences.

***Table 3.6* Location of Targeted Vocabulary**

Word	Location Title or paragraph	Main idea sentence	Detail sentence	Important for meaning of sentence or passage
Docile 8	Title	Yes	No	Yes
Multiple-choice 9	1	No	Yes	No
Tenure 9	1 (Title & paragraph 1 = 21% of words in text and contains 14% of words claimed as unknown 7-10 times.)	No	Yes	No
Constraints 7	2	No	Yes	Yes
Leeway 10	2 (Paragraph 2 = 7% of words in text and contains 9% of words claimed as unknown 7-10 times.)	No	Yes	Yes
Strive 8	3	No	Yes	Yes
Boot Camp 7	3	No	Yes	No
Rationalized 10	3	Yes	No	No
Procedures 9	3	Yes	No	No
Spontaneity 10	3	Yes	No	Yes
Docility 10	3 (Paragraph 3 = 21% of words in text and contains 27% of words claimed as unknown 7-10 times.)	Yes	No	Yes
Tyranny 10	4	No	Yes	No
Cluster 7	4	No	Yes	No
Enormous 7	4	No	Yes	No
Intensity 9	4	No	Yes	No
Submissive 8	4	Yes	No	Yes
Malleable 10	4	Yes	No	Yes
Time-consuming 8	4 (Paragraph 4 = 26% of words in text and contains 32% of words claimed as unknown 7-10 times.)	Yes	No	Yes
Equivalent 10	5	Yes	No	Yes
Omnipresent 10	5 (Paragraph 5 contains 15% of words in text and 9% of words claimed as unknown 7-10 times.)	Yes	No	No
Remedial 9	6	No	Yes	No
Tailors 7	6 (Paragraph 6 contains 10% of words in text and 9% of words claimed as unknown 7-10 times.)	No	Yes	No
Total 22		10	12	10

Hulstijn (1993) found that rather than looking up every word, students were influenced by the nature of the reading task when making decisions about what words to look up. If students recognize main ideas, they may be more likely to look up unknown words in main idea sentences than in detail sentences. Look ups may also be influenced by the importance of words to understanding the sentence or text. In Table 3.6 each word has been categorized

as either important or not to an understanding of the sentence or overall passage. While not a main focus of the present study, by asking the students to underline words looked up, it is possible to examine whether students look up every unknown word or use more selective strategies. Another possibility that was considered was only including words in the vocabulary tests from main idea sentences or words important to understanding the sentence or passage. However, because of uncertainty about whether the participants have the reading skills to only look up words in main idea sentences or whether it is possible to decide whether a word is important before knowing what it means, vocabulary from detail sentences and items that were not necessarily key to understanding the sentences or the passage were retained on the vocabulary tests.

3.5 Dependent Measures

Three dependent measures were used: a receptive vocabulary checklist (RVC), a productive vocabulary definition selection test (PVDS), and a reading comprehension test.

3.5.1 Receptive Vocabulary Checklist Test

At each of the three test administrations, the receptive vocabulary checklist test (RVC) in Appendix III was the first test. The RVC is a reduced version of the pilot checklist in Table 3.4. In the experiment however, the RVC was not used to estimate the percentage of unknown words in the text. Instead, the RVC was used as a pre and post-test for known items to measure receptive vocabulary gains from reading.

The RVC contains a total of 50 items and requires only that students indicate whether or not they know the lexical item. The RVC utilizes 20 difficult lexical items from the pilot checklist selected by 70% of the students as unknown and 20 imaginary words to provide a correction for students over claiming known words. Ten easy words that none of the students selected as unknown in the pilot were included so that students were not made uncomfortable by possibly having to select all the words as unknown. Easy words *were not counted* when scoring the RVC and did not have the same meanings as the target difficult

vocabulary words. Table 3.7 shows the different types of RVC items in the order that they appear on the test.

In section 2.7, what a word looks like and what meaning a word form signals are considered part of the receptive knowledge of a word. Checklist tests are sensitive to partial knowledge because students mark words as known even if they only partially understand the words (Anderson and Freebody, 1983 in Nagy, Herman and Anderson, 1985). Because the checklist test measures very minimal types of vocabulary knowledge, it was chosen in preference to a multiple choice receptive vocabulary test that may require more precise knowledge.

Table 3.7 **Items on Receptive Vocabulary Checklist Test**

---

abrivator bloss "boot camp" cluster constraints creative customers dinated docility education equivalent exam experienced follow glinder gulate heam independent instined intensity julique leeway malleable minurite moldarian "multiple-choice" nud omnipresent procedures pindle rationalized recund remedial revictive science servate skilled sleem snickle spontaneity strive submissive tailors technology tenure "time-consuming" tomby tripter tyranny undiration

Color code: black = imaginary item, red = difficult item, blue = easy item (not counted for scoing)

The grammatical forms of difficult words as they were used in the text were: 11 nouns 55%, 7 adjectives 36%, 2 verbs 10%. Easy words were included in roughly the same proportions: 5 nouns 50%, 4 adjectives 40%, 1 verb 10%. Grammatical categories can not be assigned to the imaginary words. However, pilot versions of "heamly" and "blossly" were changed to "heam" and "bloss" because the ly ending suggests adverbs which were not included in the difficult target vocabulary words.

Scores for the difficult lexical items and imaginary words were calculated with the formula P(k) = P(h) – P(fa)/1 – P(fa). Maera and Buxton explain,

> This formula derives directly from stimulus detection theory studies, P(h) (i.e., the probability of making a 'hit') in our study is the proportion of real words that the testee recognizes (RY); P(fa) (i.e. the probability of a 'false alarm') is the proportion of imaginary words the testee claims to know (IY).

> The formula adjusts the RY score downwards if the IY is large. P(k) in signal detection theory represents the likelihood of a real target actually being acknowledged; in our study it indicates how many of the target words the testee can be deemed to know. (Maera and Buxton, 1987:147)

The maximum possible score for known words is 20.

3.5.2 Productive Vocabulary Definition Selection Test

At each of the three test administrations, the productive vocabulary definition selection test (PVDS) in Appendix IV was the second test. The PVDS utilizes the same test form as the Receptive Vocabulary Checklist (RVC) with the addition of a list of Japanese translations for the 20 difficult words. Each translation was selected to conform as closely as possible to the meaning sense used in the original text. Students wrote the number of the English word next to the Japanese word that had the same meaning.

The test requires starting from a Japanese meaning and selecting the correct English form from the list of 50 items on the checklist, which corresponds to the productive knowledge of what word form can be used to express a particular meaning cited in section 2.7. Laufer and Goldstein consider multiple choice items that begin with an L1 meaning and require the selection of an appropriate L2 form to be tests of “Active Recognition” while items that require the testee to supply an L2 form for an L1 meaning are considered “Active Recall” (2004: 406) where the term active is equivalent to the term productive as it is used in this paper. The PVDS could have been made more clearly a measure of productive ability by only supplying the Japanese translations and requiring testees to supply the English form but this would have required recall. The PVDS relies on recognition, which is easier than recall and thereby increases the likelihood of observing small amounts of productive vocabulary learning. This format also avoids the difficulties of writing multiple choice distractors. So that the items are as independent of each other as possible and to discourage process of elimination guessing, the test instructions falsely state that some of the English words may have the same meaning and that answers may be used more than once. Each correctly answered question was given one point for a maximum of 20 points.

3.5.3 True/False Comprehension Test

The comprehension test was administered after the RVC and PVDS on the day that the students read the text. The comprehension test consists of 28 statements about the text translated into Japanese to generate items to be marked by the subjects as true or false. For example item number two, “According to the article, students have many choices in the types of course they can take” was translated to 記事に依ると、学生には様々なコースの選択があります. The items were presented in subjects’ L1 to ensure that the comprehension being measured was comprehension of the passage and not the comprehension of the test questions/input (Bachman, 1990: 127-8). An additional concern was that the comprehension questions should not exclusively target vocabulary items but should measure a broader type of comprehension. Comprehension questions can simply be vocabulary questions with context (Read, 2000: 10). The questions in the present comprehension test provide a proposition that includes the meaning of the difficult vocabulary without necessarily focusing on the meaning of the difficult vocabulary.

Referring to dictionaries or the text while taking the comprehension tests was not permitted. Parts of the text may have been understood at the moment of reading but not retained long enough to answer test questions. Thus, the reading construct being tested is actually comprehension plus short-term recall. Consideration was given to allowing access to the text while answering the questions but this might have substantially impacted the reading by transforming the task from reading for overall meaning into a search for isolated bits of information.

A pilot version of the test with 37 students from another instructor’s class with students in the same TOEFL range was conducted in which the test was taken without reading the text passage to examine the effect of guessing skills and background knowledge. To the maximum extent possible, the comprehension test should only measure comprehension and not other unwanted variables such as testwiseness (Bachman, 1990: 114) and background knowledge (ibid.: 273-4). The average was 14.89 which is very close to the midpoint of hypothetical population score of 14 that would be expected for random guessing on a 28

item true-false test. A one-sample t-test indicated that 14.89 was just barely significant at an alpha level of .05. However, there were 9 questions where 25 or more students chose the correct answer and 5 items where 12 or less students chose the correct answer. In both of these cases, the t-test result was significant and suggested that half of the test items were either positively or negatively affected by background knowledge (See Appendix V for example t-test calculations). In a comprehension test of a non-fiction expository text, it is extremely difficult to completely eliminate the effects of background knowledge. However, five of the items that received more than 70% correct responses were rewritten to reduce the likelihood of ceiling effects that would hamper finding statistically significant differences in the experiment groups. The final version of the test is in Appendix VI.

Table 3.8 itemizes each question by the target paragraph, whether the area is a main idea or an example area, and by a subjective impression of the degree of inference required to answer the question. The categorization by degree of inference is provided primarily to describe the test characteristics. However, it also permits a more detailed analysis of responses on comprehension questions should any unusual patterns of response appear. In Appendix VII, the questions are matched with the targeted text passage.

Table 3.8 **Comprehension Question Characteristics**

Q#	T	F	Area	Main	Detail	No/ Minimal Inference	Minor Inference	Inference	Pilot
7	X		1	X		X			51%
9		X	1		X		X		70%*
18		X	1		X		X		27%*
15	X		1		X		X		N/A
22		X	2	X		X			41%
2		X	2		X	X			46%
25	X		3		X	X			70%*
27	X		3	X			X		N/A
13	X		3	X		X			54%
17		X	3	X		X			68%*
24	X		3	X		X			59%
4		X	4	X		X			N/A
12	X		4		X			X	70%*
23		X	4		X	X			65%
28	X		4	X			X		N/A
21	X		4	X		X			24%*
20	X		5	X		X			59%
8		X	5		X	X			38%
26		X	5		X		X		16%*
14	X		5	X		X			65%
19		X	5	X				X	38%
3	X		6		X	X			N/A
5		X	6	X		X			24%*
6		X	Multi	X				X	35%
16		X	Multi	X				X	57%
11		X	Multi	X				X	51%
10		X	Multi	X				X	32%*
1	X		Multi	X				X	57%
To -tal	13	15		18	10	15	6	7	

Abbreviations: Q# Question number, T answer is true, F answer is false, Area shows area of text; numbers represent the paragraph while multi indicates that the question covers multiple paragraphs. Pilot shows the percentage of 37 students in a pilot test that correctly answered the question without reading the text. An asterisk indicates that the percentage is significantly above or below a hypothetical population mean of 50%. N/A indicates that more than 70% of students correctly answered the pilot version of the question and that the question was rewritten for the final version of the test. Rewritten questions were not further piloted because of a lack of additional subjects.

So that no section of the text received a disproportionate number of questions, the percentage of words in each paragraph was compared to the percentage of the 23 questions

that targeted particular paragraphs and the results are presented in Table 3.9. There were also an additional five questions that were general in nature and covered multiple sections.

Table 3.9 **Coverage of Text by Comprehension Questions**

Paragraph #	# of words in text section	% of words in text section	# of Comp Qs targeting section	% of Comp Q's targeting section
Title + 1	141	21%	4	17%
Two	45	7%	2	9%
Three	137	21%	5	22%
Four	173	26%	5	22%
Five	101	15%	5	22%
Six	66	10%	2	9%
Total	663		23	

One point was awarded for each correct answer for a maximum possible score of 28.

4 Results and Statistical Significance

Table 4.1 lists the average scores of all groups on the RVC and PVDS vocabulary measures at each administration, the average scores of all groups on the comprehension test and the average time taken to read the text for each group.

On the RVC and PVDS immediate post-tests and two-week post-tests, the ranks of the group mean scores from highest to lowest were: electronic dictionaries (ED) > printed dictionaries (PD) > no dictionaries (ND)> no text and no dictionaries (NTND). On the RVC and PVDS two-week pre-tests, there was very little difference between the respective test scores of the different groups. Over time, the ED, PD, and ND groups all had increases on both the RVC and PVDS from the two-week pre-test to the immediate post-test and decreases from the immediate post-test to the two-week post-test. The RVC and PVDS two-week post-test scores for the ED, PD, and ND groups were all higher than their respective two-week pre-test scores. The NTND group experienced a small increase in scores each time it retook the RVC. However, on the PVDS the NTND scores increased from the pre-test to the immediate post-test but did not change from the immediate post-test to the two-week post-test.

On the comprehension tests, the ranks of the group mean scores from highest to lowest were: ED>ND>PD>NTND.

In regard to reading times, the PD group took the most time to read the text followed by the ED group and ND group.

To address the research questions, the next section examines the statistical significance of the results.

Table 4.1 **Overview of Results**

	n	RVC Pre	PVDS Pre	RVC Post	PVDS Post	Comp Post	Reading Time (minutes)	RVC Delayed Post	PVDS Delayed Post
ED	53	1.91	4.43	11.60	7.75	16.40	30	6.74	6.02
PD	49	1.49	3.90	9.53	5.82	15.16	33	5.16	4.73
ND	48	1.79	4.125	4.33	4.58	15.31	18	4.33	4.50
NT ND	24	2.04	4.13	3.125	4.54	14.08	N/A	4.125	4.54

Abbreviations: ED = Electronic Dictionary Group, PD = Printed Dictionary Group, ND = No dictionary Group, NTND = No Text No Dictionary Group, n = number of participants in group, RVC = Receptive Vocabulary Checklist, PVDS = Productive Vocabulary Definition Selection Test, Comp = Comprehension Test, Pre = Two-Week Pre-test, Post = Immediate Post-experiment test, Delayed Post = Two-Week Post-test, N/A = not applicable (NTND group did not read text)

4.1 Statistical Significance of RVC and PVDS Vocabulary Measures

The first research question is whether there are significant differences in pre and post-reading receptive and productive vocabulary measures between the exposure to text only, exposure to text and PD, exposure to text and ED, and no exposure to text or dictionary groups. Using ANOVA this section first examines the RVC and then the PVDS test scores of different groups during the same test administrations and each individual group's scores over time. Six one way analysis of variance (ANOVA) calculations were used to compare the RVC and PVDS scores of different groups at each test administration. Eight ANOVA calculations were done to determine significant differences on the RVC and PVDS scores over time for each group. ANOVA calculations were done with Excel with manual calculation of the Scheffe tests. The alpha level for all ANOVA calculations was set at .05. While the calculation of multiple one way ANOVA increases the likelihood that one or more of the probability based calculations will be in error, I did not have access to superior statistical software that would have permitted more sophisticated statistical procedures such as MANOVA. Although ANOVA is not ideal, it is adequate for the purposes of this experiment. Table 4.2 is an excerpt of Table 4.1 showing only the RVC scores for ease of reference.

Table 4.2 **Overview of RVC Results**

Group	RVC Pre	RVC Post	RVC Delayed Post
ED	1.91	11.60	6.74
PD	1.49	9.53	5.16
ND	1.79	4.33	4.33
NTND	2.04	3.125	4.125

On the RVC two-week pre-test, an ANOVA analysis did not reveal any significant differences $F(3,170) = .593$, $p = .62$) between the ED, PD, ND, and NDNT groups.

Of primary interest for answering the first research question with regard to the RVC is a comparison of the immediate post-test scores for each group. The ANOVA yielded $F(3,170) = 33.615$, $p < .05$) indicating that there was a significant difference among the four groups on the immediate post-test RVC. To determine which groups had significantly different scores, Post Hoc Scheffe tests for each pair were calculated and are reported in Table 4.3. The alpha level for all Scheffe tests in this study is .05 so in all instances where significance is indicated, p is less than .05. The Scheffe tests show that on the immediate post-test RVC scores there was no significant difference between the ED and PD groups, but that the ED group's score was significantly higher than the ND and the NTND scores. The PD group was also significantly higher than the ND and NTND groups. There was however no significant difference between the ND and NTND groups.

Table 4.3 **Scheffe Test Values**: **RVC Immediate Post-test of All Groups**

Immediate post-test	ED RVC	PD RVC	ND RVC	NTND RVC
ED RVC	N/A	1.83	*22.22	*19.81
PD RVC	X	N/A	*10.93	*11.03
ND RVC	X	X	N/A	.39
NTND RVC	X	X	X	N/A

* = Significant Result

The next aspect of the first research question is how each group performed relative to the other groups on the RVC two-week post-test. The ANOVA yielded $F(3,170) = 4.465$, $p < .05$) indicating that there was a significant difference among the four groups on the two-

week post-test RVC. Post Hoc Scheffe tests are reported in Table 4.4. The Scheffe tests indicated that on the two-week post-test RVC scores there was no significant difference between the ED and PD groups, but that the ED group's score was significantly higher than the ND and the NTND scores. There were no significant differences between any of the other group pairs.

Table 4.4 **Scheffe Test Values**: **RVC Two-Week Post-test of All Groups**

Two-week post-test	ED RVC	PD RVC	ND RVC	NTND RVC
ED RVC	N/A	1.49	*3.45	*2.67
PD RVC	X	N/A	.40	.41
ND RVC	X	X	N/A	.02
NTND RVC	X	X	X	N/A

* = Significant Result

The final set of ANOVA and Scheffe tests on the RVC check for significant changes in scores over time for each group. These are reported in Table 4.5. The Post Hoc Scheffe Test pairings are reported in the last three columns of the table.

The ANOVA comparison of the ED group's two-week pre-test, immediate post-test and two-week post-test RVC scores indicated that there was a significant difference $F(2,156) = 74.386$, $p<.05$) in scores among the three testings. The Scheffe tests indicated a significant increase from the pre-test to the immediate post-test, a significant decrease from the immediate post-test to the two-week post-test but still showed a significant increase from the two-week pre-test to the two-week post-test.

The ANOVA comparison of the PD group's two-week pre-test, immediate post-test and two-week post-test RVC scores was significant $F(2,144) = 53.399$, $p<.05$). The Scheffe tests indicated a significant increase from the pre-test to the immediate post-test, a significant decrease from the immediate post-test to the two-week post-test but still showed a significant increase from the two-week pre-test to the two-week post-test.

The ANOVA comparison of the ND group's two-week pre-test, immediate post-test and two-week post-test RVC scores was significant $F(2,141) = 12.659$, $p<.05$). The Scheffe tests indicated a significant increase from the pre-test to the immediate post-test, no significant change from the immediate post-test to the two-week post-test and a significant increase between the two-week pre-test and two-week post-test scores.

The ANOVA comparison of the NTND group's two-week pre-test, immediate post-test and two-week post-test RVC scores was significant $F(2,69) = 3.538$, $p<.05$). Scheffe tests indicated no significant change from the pre-test to the immediate post-test, or from the immediate post-test to the two-week post-test but showed a significant increase between two-week pre-test and two-week post-test scores.

Table 4.5 **RVC Over Time of ED, PD, ND and NTND Groups: ANOVA and Scheffe Test Values**

	Degrees of Freedom	ANOVA Value of F	Critical Value of F at p = .05	Two-week Pre-test Versus Immediate Post-test	Immediate Post-test Versus Two Week Post-test	Two Week Pre-test Versus Two Week Post-Test
ED	(2,156)	*74.386	3.054	*74.38	*18.74	*18.45
PD	(2,144)	*53.399	3.059	*53.26	*15.71	*11.12
ND	(2,141)	*12.659	3.060	*9.50	0	*9.50
NTND	(2,69)	*3.538	3.130	.956	.815	*3.535

* = Significant Result

The remainder of this section reports the ANOVA results for the PVDS. Table 4.6 is an excerpt of Table 4.1 showing only the PVDS scores.

Table 4.6 **Overview of PVDS Results**

Group	PVDS Pre	PVDS Post	PVDS Delayed Post
ED	4.43	7.75	6.02
PD	3.90	5.82	4.73
ND	4.125	4.58	4.50
NTND	4.13	4.54	4.54

On the PVDS two-week pre-test, an ANOVA analysis did not reveal any significant differences $F(3,170) = .755$, $p = .52$) between the ED, PD, ND, and NDNT groups.

A comparison of the immediate post-test scores for each group is of primary interest for answering the first research question with regard to the PVDS. The ANOVA yielded $F(3,170) = 14.587$, $p<.05$) indicating that there was a significant difference among the four groups. Post Hoc Scheffe tests are reported in Table 4.7. The ED group's scores were significantly higher than the PD, ND and NTND scores. The PD group's scores were not significantly higher than the ND and NTND scores. There was also no significant difference between the ND and NTND groups.

Table 4.7 **Scheffe Test Values**: **PVDS Immediate Post-test of All Groups**

Immediate post-test	ED PVDS	PD PVDS	ND PVDS	NTND PVDS
ED PVDS	N/A	*4.50	*11.92	*8.02
PD PVDS	X	N/A	1.73	1.20
ND PVDS	X	X	N/A	.03
NTND PVDS	X	X	X	N/A

* = Significant Result

Another important aspect of the first research question is how each group performed relative to the other groups on the PVDS two-week post-test. The ANOVA yielded $F(3,170) = 4.626$, $p<.05$) indicating that there was a significant difference among the four groups. Post Hoc Scheffe tests reported in Table 4.8 show that there was no significant difference between the ED and PD groups, but that the ED group's score was significantly higher than the ND group's score. There were no significant differences between any of the other group pairs.

Table 4.8 **Scheffe Test Values**: **PVDS Two-Week Post-test of All Groups**

Two-week post-test	ED PVDS	PD PVDS	ND PVDS	NTND PVDS
ED PVDS	N/A	2.58	*3.57	2.21
PD PVDS	X	N/A	.08	.04
ND PVDS	X	X	N/A	.002
NTND PVDS	X	X	X	N/A

* = Significant Result

The final set of ANOVA and Scheffe tests on the PVDS check for significant changes in scores over time for each group. These are reported in Table 4.9. The Post Hoc Scheffe Test pairings are reported in the last three columns of the table.

The ANOVA comparison of the ED group's two-week pre-test, immediate post-test and two-week post-test PVDS scores indicated a significant difference $F(2,156) = 20.296$, $p<.05$) in scores among the three testings. The Scheffe tests showed a significant increase from the pre-test to the immediate post-test, a significant decrease from the immediate post-test to the two-week post-test but still showed a significant increase from the two-week pre-test to the two-week post-test.

The ANOVA comparison of the PD group's two-week pre-test, immediate post-test and two-week post-test PVDS scores was significant $F(2,144) = 8.18$, $p<.05$) . The Scheffe tests indicated a significant increase from the pre-test to the immediate post-test. Scores decreased from the immediate post-test to the two-week-post test but the decrease was not significant. There was a slight increase from the two-week pre-test to the two-week post-test but the increase was not significant.

The ANOVA comparison of the ND group's two-week pre-test, immediate post-test and two-week post-test PVDS scores showed no significant difference $F(2,141) = .646$, $p = .53$) in scores among the three testings.

The ANOVA comparison of the NTND group's two-week pre-test, immediate post-test and two-week post-test PVDS scores was not significant $F(2,69) = .692$, $p = .50$).

Table 4.9 **PVDS Over Time of ED, PD, ND and NTND Groups: ANOVA and Scheffe Test Values**

	Degrees of Freedom	ANOVA Value of F	Critical Value of F, p =.05	Two week Pre-test Versus Immediate Post-test	Immediate Post-test Versus Two Week Post-test	Two Week Pre-test Versus Two Week Post-Test
ED	(2,156)	*20.296	3.054	*20.30	*5.55	*4.62
PD	(2,144)	*8.186	3.059	*8.14	2.59	1.55
ND	(2,141)	.646	3.060	N/A	N/A	N/A
NTND	(2,69)	.692	3.130	N/A	N/A	N/A

* = Significant Result

4.2 Statistical Significance of Comprehension Scores

The second research question is whether there are there significant differences in reading comprehension scores between the exposure to text only, exposure to text and PD, exposure to text and ED and no exposure to text or dictionary groups. Table 4.10 is an excerpt of Table 4.1 showing only the comprehension test scores.

Table 4.10 **Comprehension Test Scores**

Group	Comp Post
ED	16.40
PD	15.16
ND	15.31
NTND	14.08

The ANOVA comparison of the comprehension scores of the ED, PD, ND and NDNT groups yielded $F(3,170) = 3.022$, $p<.05$) indicating that there was a significant difference among the four groups. The Scheffe tests reported in Table 4.11 indicate that the ED group was significantly higher than the NTND group. However, there were no significant differences between any of the other group pairings.

Table 4.11 **Scheffe Test Values: Comprehension Test of All Groups**

	ED Comp	PD Comp	ND Comp	NTND Comp
ED Comp	N/A	1.20	0.92	*2.75
PD Comp	X	N/A	0.02	0.58
ND Comp	X	X	N/A	0.75
NTND Comp	X	X	X	N/A

* = Significant Result

4.3 Statistical Significance of Reading Time Scores

The third research question is whether there are there significant differences in the time taken to read the text between the exposure to text only, exposure to text and PD, and exposure to text and ED groups. Table 4.12 is an excerpt of Table 4.1 showing only the reading times of each group.

Table 4.12 **Reading Time**

Group	Reading Time (minutes)
ED	30
PD	33
ND	18
NTND	N/A

The ANOVA comparison of the reading times of the ED, PD, and ND groups yielded $F(2,147) = 51.060$, $p<.05$) indicating that there was a significant difference in reading times among the three groups. The Scheffe tests reported in Table 4.13 show that there was no significant difference between ED and PD group reading times but that both the ED and PD groups spent significantly more time reading than the ND group.

Table 4.13 **Scheffe Test Values: Reading Times of ED, PD, and ND Groups**

	ED Reading Time	PD Reading Time	ND Reading Time
ED Reading Time	N/A	1.61	*31.03
PD Reading Time	X	N/A	*44.97
ND Reading Time	X	X	N/A

* = Significant Result

5 Discussion: Effects of Dictionary Use on Vocabulary Learning, Comprehension and Reading Time

5.1 Effects of Dictionary Use on Vocabulary Learning

The first issue is whether the subjects learned targeted vocabulary as the result of reading the text. The ND group's scores on both the immediate post-test and two-week post-test showed no significant differences with the NTND group on either the RVC or the PVDS. This suggests that reading the text without dictionaries did not result in improved recognition or partial knowledge of the targeted vocabulary. This finding does not contradict Nagy, Herman, and Anderson (1985), Nagy, Anderson, and Herman (1987), or Nagy (1997) that showed that students can learn vocabulary from context but confirms that for vocabulary learning to take place from reading that the texts used are of central importance. The text used in the experiment was probably too difficult for the subjects and likely contained too many unknown words for the students to make enough sense of the text for learning from context to take place (Bensoussan, Sim, and Weiss, 1984, Hu Hsueh-chao and Nation, 2000).

In contrast, both the ED (11.60) and PD (9.53) groups scored significantly higher than both the ND (4.33) and NTND (3.125) groups on the RVC immediate post-test. This finding suggests that dictionary use while reading the text contributed more to improved recognition and partial receptive knowledge of the targeted vocabulary than reading the text without a dictionary. However, these gains proved highly perishable. Both the ED and PD groups experienced significant declines from the immediate post-test (11.60, 9.53) to the two-week post-test (6.74, 5.16).

On the PVDS immediate post-test, the ED group (7.75) scored significantly higher than both the ND (4.58) and NTND (4.54) groups. This finding suggests that electronic dictionary use contributed more to a limited productive knowledge of the targeted vocabulary than reading the text without a dictionary. As was the case with the RVC, the ED group's PVDS score declined significantly from the immediate post-test (7.75) to the

two-week post-test (6.02). On the other hand, the PD group's PVDS scores at the immediate post-test and at the two-week post-test (5.82, 4.73) were not significantly different from either the ND (4.58, 4.50) or NDNT (4.54, 4.54) groups. Thus reading the text with a printed dictionary does not appear to have contributed more to productive knowledge of the targeted vocabulary than reading the text without a dictionary. To the extent that dictionary use sometimes resulted in more vocabulary knowledge than reading the text without a dictionary, the findings tend to be consistent with Knight's (1994) and Luppescu and Day's (1993) studies.

The next issue of interest is which type of dictionary aided vocabulary learning the most. On the RVC, comparisons at the immediate post-test and two-week post-test of the ED scores (11.60, 6.74) with the PD group's scores (9.53, 5.16) were not significant reflecting a relatively small difference in scores which on first inspection would seem to suggest that there was no difference between using electronic or printed dictionaries. However, the mean scores of the ED group were higher than the PD group's scores. At the RVC two-week post-test, the higher score for the ED group (6.74) resulted in a significant difference when compared to the ND (4.33) and NTND (4.125) groups but the lower score of the PD group (5.16) resulted in non-significant results suggesting that electronic dictionary use contributed more to retaining recognition and partial receptive knowledge of the targeted vocabulary than printed dictionary use. This result runs counter to Stirling's (2003) and Koyama and Takeuchi's (2004a) claim that alphabetical look up procedures promote deeper processing leading to better retention than the key punch look up of electronic dictionaries.

On the PVDS immediate post-test, the ED group's score (7.75) was significantly higher than the PD group's score (5.82). This finding suggests that electronic dictionaries contributed more to productive knowledge than printed dictionaries. At the PVDS two-week post-test, the ED group's score (6.02) was higher than the PD group's (4.73) but there was no significant difference between the scores. However, the slightly higher ED score resulted in a significant difference between the scores of the ED group (6.02) and the ND group (4.50) that was not evident between the PD group (4.73) and the ND (4.50) and

NDNT (4.54) groups at the two-week post-test. Oddly however, the ED group's score (6.02) was not significantly different from the NTND group (4.54) at the two-week post-test. This was the result of the small number of subjects in the NTND group affecting the calculation. If the NTND group had had a larger number of subjects with the same average score (4.54), the result would have been significant. Consequently even at the two-week post-test, it appears that electronic dictionary use contributed more to productive knowledge of the target vocabulary than printed dictionary use. It is also important to note that of all the groups, only the ED group had a significant increase from its two-week pre-test to the two-week post-test PVDS scores.

The fact that electronic dictionary use seems to have resulted in more receptive knowledge at the two-week post-test and more productive vocabulary knowledge at both the immediate and two-week post-tests than printed dictionary use is a novel finding. For instructors not familiar with electronic dictionaries this study suggests that electronic dictionaries should not be regarded with suspicion or as in some way inferior to printed dictionaries. In fact, for retention of new words electronic dictionaries may actually be superior to printed dictionaries. Why this would be the case is not immediately apparent. The current generation electronic dictionaries have large memories that usually contain near mirror images of the entries found in printed dictionaries. In fact, some electronic dictionaries simply contain digital versions of particular print dictionaries. One possibility is that regular use of electronic dictionaries has made students less proficient with printed dictionaries. A survey of 219 of my EFL university students revealed that 95% own ED and that of ED owners 97% bring them to class regularly. While 81% own printed English to Japanese dictionaries, only 3% bring them to class. Another possible reason considered for differences in the performance between the ED and PD groups was that perhaps the groups had different ability levels. However, the data from the experiment does not appear to indicate that there were major differences in the starting knowledge of the target words as evidenced by the fact that there were no significant differences between groups on the RVC or PVDS scores on the two-week pre-tests. Additionally, as shown in Table 5.1, there were no significant differences $F(3,170) = 0.901$, $p = .44$) between groups on the Nation 3000

Test which is a more general indication of vocabulary knowledge within the students' ability level.

Table 5.1 **Group Means on Nation 3000 Test**

ED Group n=53	PD Group n=49	ND Group n= 47	NTND Group n=24
21.92	20.94	21.33	22.33

One point overlooked during the design stage was that many electronic dictionaries have a history function that can record words looked up for subsequent access. Use of this function could have given the ED group an advantage on the two-week post-tests. The only way to absolutely control for this would be for the researcher to supply the dictionaries. Even so, in informal conversations with about thirty students none seemed familiar with this function. One final possibility identified and further investigated in section 5.2 is that an increased look-up speed in electronic dictionaries results in more vocabulary knowledge.

From a pedagogical perspective, the declines on the vocabulary measures from the immediate post-test to the two-week post-test are a matter of some concern. The targeted words were probably not sufficiently common enough to be re-encountered between tests. While satisfactory for control purposes of the experiment, single encounters with a word are likely to be forgotten in a relatively short period of time, which argues in favor Nation's (2001) position for repetition of newly introduced words in pedagogic materials so that words looked up have a better chance of being retained.

Another issue of pedagogical interest is whether word focused learning activities are a more efficient means of learning vocabulary than reading or dictionary use. On the RVC, the NTND group was able to gain about two (2.085) words from the pre-test to the two-week post-test. This gain appears to be due to students learning words from translations on the PVDS test. For instance, in the case of "Multiple-choice", a sample of 79 pre-tests revealed that only 41% of the students had checked the item as known but on the PVDS 92% of the students were able to make a correct match. Even though the meaning of "multiple" was probably not initially known to most students, they probably knew the meaning of "choice"

and used this information to make a match on the PVDS which resulted in higher pre-test scores on the PVDS than on the RVC and a small but significant increase of the NTND group's RVC scores during the post-tests. This process of learning from the tests could be considered a type of word focused learning activity. The figures in Table 5.2 assume that like the NTND group that the other groups gained 2.085 words attributable to learning from taking the tests and subtracts this from the gains made from the two-week pre-test to the two-week post-test. Although it was not the intention that the tests serve as a word focused learning activity, the NTND group's gain of 2.085 words was superior to the adjusted gains of groups reading the text without a dictionary or reading the text with printed dictionaries. This suggests that where the sole concern is vocabulary learning, that activities actually designed for the purpose of word focused learning are probably more efficient than reading or dictionary use. The relatively small vocabulary gains from reading in this experiment also tend to confirm Hill and Laufer's (2003) argument that L2 learners cannot be expected to learn vocabulary exclusively through extensive reading. Before students reach a level of vocabulary knowledge where learning vocabulary through pleasure reading is possible, word focused activities should be included as an important part of the curriculum.

Table 5.2 **RVC Pre-test to Two-Week Post-test Gains Adjusted for Possible Learning from Tests**

ED RVC	PD RVC	ND RVC	NTND RVC
2.745	1.585	0.455	0

5.2 Effects of Dictionary Use on Comprehension

On the comprehension measure, the ED group's scores (16.40) were low but significantly higher than the scores of students who had not read the text (14.08) suggesting that electronic dictionaries were only just marginally helpful in aiding comprehension. Because the students who had read the text either with printed dictionaries (15.16) or without dictionaries (15.31) did not score significantly better on the Scheffe test than the students who had not read the text, it would appear that printed dictionaries did not improve performance on the comprehension test and that both the PD and ND groups on average may not have understood the text. This finding is consistent with studies such as

Bensoussan, Sim and Weiss (1984), Koyama and Takeuchi (2004b) and Albus, Thurlow, Liu, and Bielinski (2005) that have had difficulties establishing that using dictionary usage consistently improves reading comprehension.

As an alternative to ANOVA and Scheffe test comparisons, the one-sample t-test described in section 4.6 was used to determine whether the ED, PD, and ND groups were significantly different from a random guessing population with a score of 14. By plugging the critical value of t for the NTND into the equation for the one-sample t-test, it was estimated that a score greater than or equal to 15.1 on the comprehension score would be significant when compared to a random guessing population score of 14. Using this method, the ED, PD and ND scores are just barely significantly higher than what would be predicted by random guessing.

Only examining the mean scores conceals some differences between groups that became apparent by counting the number of students that were able to achieve scores of 71% or more (20 or more correct out of 28). The results of this tally are reported in Table 5.3. The number of students in each group that achieved a satisfactory level of comprehension suggests that not only electronic dictionaries but also printed dictionaries might be more helpful in aiding comprehension than inferencing strategies alone which tends to confirm Knight's (1994) finding that dictionary usage aids comprehension. With a less difficult text or with students of a higher ability level, larger differences between the ND and NTND groups might have also been apparent. Even so, it does not appear that as a general policy that discouraging the use of dictionaries in favor of contextual guessing is justified since in some instances dictionary use permits bootstrapping to a level of comprehension not possible without looking up words.

Table 5.3 **Percentage of Students with Adequate Comprehension for All Groups**

ED Comp 20 plus	PD Comp 20 plus	ND Comp 20 plus	NTND 20 plus
10 (18.9%)	8 (16.3%)	3 (6.25%)	1 (4.17%)

5.3 Effects of Dictionary Use on Reading Time and Other Findings

There was no significant difference between reading times of the ED and PD groups. There were however very significant differences between the ED group that took 66% more time to read than the ND group and the PD group that took 83% more time than the ND group. The finding that dictionary use resulted in increased reading time is consistent with Knight (1994) where dictionary use resulted in 42% more time to read the text. While not significant, printed dictionary users spent 10% more time reading than the electronic dictionary users. This is similar to Koyama and Takeuchi's (2004b) finding that printed dictionary users spent 16% more time reading than electronic dictionary users.

The extra time required to read by the dictionary groups was clearly spent looking up unknown words. To compare differences between the number of words looked up by the ED and PD groups, words looked up and circled by students during the reading were tallied and compared with the independent t test, which was used instead of ANOVA because only two groups were being compared. The results are reported in Table 5.4. The ED group looked up significantly more words (11) targeted in the vocabulary tests than the PD group (8). Interestingly, both groups did not report looking up a substantial proportion of the targeted unknown words. Because the look up counts relied on student self-reporting, some students may not have indicated all words looked up. Another possibility however is that students exercised some type of strategic decision making process when looking up words (see section 3.4.2). Unfortunately, investigating this involves a complex item by item analysis, which is not possible within the scope of this paper. When considering all words looked up, the ED group (26) was again significantly higher than the PD group (19), which is consistent with Weschler and Pitts (2001) finding that pocket electronic dictionaries enabled students to look up more words than conventional dictionaries. While the difference in reading time between the dictionary groups was not significant on the ANOVA, on the independent t-test the ED group looked up significantly more words per minute (.86) than the PD group (.58). It is probable that the speed of looking up words with an electronic dictionary and the resulting increase in the number of words looked up was a factor in cases where the ED group performed better than the PD group on the experimental

measures. Additionally, it does not appear that looking up more words interferes with comprehension. The ED group looked up the most words and had the highest average comprehension scores.

Given that both the ED and PD groups spent considerably more time reading than the ND group, it is understandable why some educators might conclude that students overuse dictionaries. However, because even a small amount of unknown vocabulary may adversely affect comprehension, it is possible that students "over using" dictionaries is not evidence of a fault of the students but is in fact evidence that the teacher has misjudged the difficulty of a text that contains too much unknown vocabulary. By observing the amount of time students spend using dictionaries while reading, it may be possible to either confirm a teacher's intuition about the appropriateness of the text or to detect excessive unknown vocabulary.

Table 5.4 **Comparison of Look ups between ED and PD groups**

	ED	PD	Degrees of Freedom	Critical value of t at p = .05, two-tailed	Value of t on independent t-test
Targeted words looked up (maximum 20 possible)	11	8	100	1.98	*3.29
All words looked up (Approximately 298 Types possible: plurals, conjugations of verbs, and closely related words treated as one type)	26	19	100	1.98	*2.76
Number of all words looked up per minute	.86	.58	100	1.98	*4.36

* = Significant Result

6 Conclusion

The results of this experiment tend to support Knight's (1994) finding that dictionary usage while reading results in more vocabulary gains and improved comprehension than reading without a dictionary. Both electronic and printed dictionary users had significantly higher scores than students that read without dictionaries on the receptive vocabulary measure on the immediate post-test. Additionally, electronic dictionary users were able to achieve significantly higher scores than students that did not use dictionaries on the receptive measure at the two-week post-test and on the productive vocabulary measure at both the immediate and two-week post-tests. On the comprehension measure both printed and electronic dictionaries resulted in a larger percentage of students achieving 70% or more correct than students that read the text without dictionaries.

Students using electronic dictionaries had the highest scores on all post-reading measures. Electronic dictionary users scored significantly higher than printed dictionary users on the immediate post-test productive measure. On the two-week post-tests for both the receptive and productive measures, direct comparison of the ED and PD groups was not significant but the higher scores of the ED group were significant when compared with the students in the ND and or NTND groups whereas the PD group's scores were not significantly different than the ND and NTND scores. According to the ANOVA analysis on the comprehension test, only the electronic dictionary group's score was significantly higher than the group that did not read the text or use dictionaries. Thus, it appears that electronic dictionary usage aids vocabulary learning and comprehension more than printed dictionaries. Possibly, this is a consequence of electronic dictionary users looking up significantly more words than printed dictionary users.

However, because of the difficulty of the text and the presence of a considerable amount of unknown vocabulary, the gains on the vocabulary measures and the comprehension text were small and came at the expense of dictionary users spending considerably more time on the reading task. The text in this experiment was so difficult that the vocabulary scores and comprehension scores of students reading without dictionaries were on the whole almost

the same as students that had not read the texts. With an easier text or higher level students, it is likely that all groups would have performed better on the vocabulary and comprehension measures and that dictionary users would have spent less time looking up words. Even so, by extrapolating the results of this study, it appears probable that dictionary users would still have an advantage over students reading without dictionaries and that electronic dictionaries might be more helpful than printed dictionaries. To confirm this there is still, however, a need to repeat a similar experiment with a text more suited to the level of the subjects and modifications to weak points in the design, procedures and materials.

Appendix I Article and Reading Instructions to Students

Name: Student Number:

Please read and make your best attempt to understand the meaning of the article. After you finish reading, you will be tested on your reading comprehension. If you are using a dictionary, you may use your dictionary while reading but not during the tests. Please underline or circle any words in the text that you look up in a dictionary. When you finish reading, raise your hand. The teacher will collect your article and give you the first test.

次の記事を読んで下さい。読み終わったら読解力のテストをします。辞書の使用を許可された人は記事を読むときは使用しても構いませんがテスト中は使用出来ません。また、辞書を使用した場合は調べた単語に〇又は下線をつけてください。読み終わったら手を上げて下さい。教師が記事を回収し、最初のテストを渡します。

Education: Learning to be docile

Universities have developed a variety of nonhuman technologies to exert control over the process of education. For instance, professors must follow certain rules and regulations. Class periods are set by the university and determine when classes must end. Students leave at the assigned time no matter where the professor happens to be in the lecture. So, too, because the university requires grading, the professor must test students. In some universities, final grades must be submitted within forty eight-hours of the final exam, which may force professors to employ computer-graded multiple-choice exams. Required evaluations by students may force professors to teach in a way that will lead to high ratings. The publishing demands of the tenure and promotion system may force professors to devote far less time to their teaching than they, and their students, would like.

Of course, students are even more controlled than professors by the university system. For example, besides the constraints already mentioned, universities often give students little leeway in the courses they may take. The courses themselves, often highly structured, force students to perform in specific ways.

Control over the university's "customers"—that is the students—actually begins long before they enter university. Grade schools in particular have developed many technologies to control students. Many schools strive, right from the start, to have students conform to their rules. Kindergarten has been described as educational "boot camp." Those who conform to the rules are thought of as good students, while those who don't are labeled bad students. Thus, as a general rule, the students who end up in college are the ones who have successfully submitted to the control mechanisms. Students are taught not only to obey authority, but also to embrace the rationalized procedures of rote learning and objective testing. More important, spontaneity and creativity tend not to be rewarded, and may even be discouraged, leading to what one expert calls "education for docility."

The clock and the lesson plan also exert control over students, especially in grade school and high school. Because of the "tyranny of the clock," a class must last until, and end at, the sound of the bell, even though learning does not often conform to the clock. Thus, even if students are just about to comprehend something, the lesson must end and the class must move on to something else. Because of the "tyranny of the lesson plan," a class must focus on what the plan requires for the day, no matter what the class (and perhaps the teacher) may find interesting. There is the example of a teacher "who sees a cluster of excited children examining a turtle with enormous fascination and intensity. Now children, put away the turtle, the teacher insists. We're going to have our science lesson. The lesson is on crabs." Overall, the emphasis tends to be on producing submissive, malleable students; creative, independent students are often, from the educational system's point of view, "messy, expensive, and time-consuming."

An even more extreme version of this emphasis appears in the child-care equivalent of the fast-food restaurant, Kinder-care. Kinder-care tends to hire short-term employees with little or no training in child care. What these employees do in the "classroom" is largely determined by an instruction book with a ready-made curriculum. Staff members open the manual to find activities spelled out in detail for each day. Clearly, a skilled, experienced, and creative teacher is not the kind of person "McChild" care centers seek to hire. Rather, relatively untrained employees are more easily controlled by the nonhuman technology of the omnipresent "instruction book."

Another example of this is the franchised Sylvan Learning Center, labeled the "MacDonald's of education." Sylvan learning centers are after-school centers for remedial education. The corporation "trains staff and tailors a McDonald's type uniformity, down to the U-shaped tables at which instructors work with their charges." Through its training methods, rules, and technologies, for-profit systems like the Sylvan Learning Center exert great control over their "teachers."

Notes: (students did not receive this information)
The McDonaldization of Society revised edition George Ritzer 1996 Pine Forge Press Thousand Oaks California USA pp106-7
Words 663
Characters excluding spaces 3610
Characters/word: 5.4

Appendix II Visual Representation of Location of Unknown Vocabulary, Main Idea and Detail Sentences in Text

Words indicated in red were chosen by at least one student in the pilot group of ten as unknown. The number of students that claimed the word as unknown is indicated by the number to the right of the word and is further coded by type size coded as follows: 1 unknown = 10 pt., 2-3 unknown = 12 pt, 4-6 unknown = 14 pt., 7-10 unknown = 16 pt.

Main ideas: general statements, main arguments, conclusions and the title are indicated by italics. Detail sentences containing mainly illustrative examples are left in non-italic text.

Education: Learning to be docile8

Universities have developed a variety of nonhuman1 technologies to exert6 control over the process of education. For instance3, professors must follow certain rules and regulations3. Class periods are set by the university and determine when classes must end. Students leave at the assigned2 time no matter where the professor happens to be in the lecture1. So, too, because the university requires grading,1 the professor must test students. In some universities, final grades must be submitted4 within forty eight-hours of the final exam, which may force professors to employ computer-graded3 multiple-choice9 exams. Required evaluations5 by students may force professors to teach in a way that will lead to high ratings5. The publishing demands2 of the tenure9 and promotion1 system may force professors to devote2 far less time to their teaching than they, and their students, would like.

Of course, students are even more controlled than professors by the university system. For example, besides the constraints7 already mentioned1, universities often give students little leeway10 in the courses they may take. The courses themselves, often highly structured5, force students to perform in specific2 ways.

Control over the university's "customers"—that is the students—actually1 begins long before they enter university. Grade schools2 in particular3 have developed many technologies to control students. Many schools strive8, right from the start, to have students conform3 to their rules. Kindergarten5 has been described1 as educational "boot camp.7" Those who conform3 to the rules are thought of as good students, while those who don't are labeled4 bad students. *Thus5, as a general rule, the students who end up6 in college are the ones who have successfully submitted4 to the control mechanisms4. Students are taught not only to obey4 authority6, but also to embrace6 the rationalized10 procedures9 of rote5 learning and objective2 testing. More important, spontaneity10 and creativity tend1 not to be rewarded6, and may even be discouraged3, leading to what one expert1 calls "education for docility10."*

The clock and the lesson plan also exert6 control over students, especially in grade school and high school. Because of the "tyranny10 of the clock," a class must last until, and end at, the sound of the bell, even though

learning does not often conform3 to the clock. Thus5, even if students are just about to comprehend something, the lesson must end and the class must move on to something else. Because of the "tyranny10 of the lesson plan," a class must focus2 on what the plan requires for the day, no matter1 what the class (and perhaps the teacher) may find interesting. There is the example of a teacher "who sees a cluster7 of excited children examining1 a turtle1 with enormous7 fascination5 and intensity9. Now children, put away the turtle, the teacher insists1. We're going to have our science lesson. The lesson is on crabs4." *Overall4, the emphasis5 tends1 to be on producing submissive8, malleable10 students; creative, independent students are often, from the educational system's point of view1, "messy5, expensive, and time-consuming8."*

An even more extreme3 version of this emphasis5 appears in the child-care1 equivalent10 of the fast-food restaurant, Kinder-care. Kinder-care tends1 to hire1 short-term1 employees1 with little or no training in child care1. What these employees1 do in the "classroom" is largely determined1 by an instruction1 book with a ready-made6 curriculum3. Staff members open the manual to find activities spelled out5 in detail1 for each day. *Clearly, a skilled, experienced, and creative teacher is not the kind of person "McChild6" care centers seek to hire1. Rather, relatively1 untrained2 employees1 are more easily controlled by the nonhuman1 technology of the omnipresent10 "instruction1 book."*

Another example of this is the franchised6 Sylvan5 Learning Center, labeled4 the "MacDonald's of education." Sylvan5 learning centers are after-school centers for remedial9 education. The corporation1 "trains1 staff and tailors7 a McDonald's type uniformity3, down to the U-shaped6 tables at which instructors work with their charges1." *Through its training methods1, rules, and technologies, for-profit6 systems like the Sylvan5 Learning Center exert6 great control over their "teachers."*

Appendix III Receptive Vocabulary Checklist Test Form (RVC)

For each item number:
Mark Bubble [1] for a word that you KNOW
Mark Bubble [2] for a word that you DO NOT KNOW
次の英単語で知っているものには（１）、知らないものには（2）をマークシートに印をつけなさい。

1. abrivator 2. bloss 3. “boot camp” 4. cluster 5. constraints

6. creative 7. customers 8. dinated 9. docility 10. education

11. equivalent 12. exam 13. experienced 14. follow 15. glinder

16. gulate 17. heam 18.independent 19. instined 20. intensity

21. julique 22. leeway 23.malleable 24.minurite 25. moldarian

26. “multiple-choice” 27. nud 28. .omnipresent 29. procedures 30. pindle

31. rationalized 32. .recund 33. remedial 34. revictive 35. science

36. servate 37. skilled 38. sleem 39. snickle 40. spontaneity

41. strive 42. .submissive 43. tailors 44. technology 45. tenure

46. “time-consuming” 47. tomby 48. tripter 49. tyranny 50. undiration

---

When you finish this part of the test, please raise your hand so that you can receive the next part of the test.
テストが終わったら手を上げてください。次のテストを渡します。

Appendix IV Productive Vocabulary Definition Selection Test (PVDS)

Name: ID number:

Choose and write the number of the English word next to the Japanese word that has the same meaning. The English words have been alphabetized to make them easier to find. Some of the English words may have the same meaning so you may choose the same answer more than once but may only have one answer for each test item. When you finish this part of the test, please raise your hand so that you can receive the next part of the test.
次の 1～20 の言葉の意味を 1～50 の英単語から選び、その番号を記入しなさい。
英単語はアルファベット順に並んでいます。同じ意味の単語があるかもしれないので、同じ答え使ってもいいが、答えは 1 つだけ選びなさい。テストが終わったら手を上げてください。次のテストを渡します。

１手続き_____２努力する_____ ３偏在する_____４自発性_____
５仕立てる_____６展性に富む_____７強度_____
８余裕_____９手間がかかる_____１０制限_____１１長期雇用契約_____
１２圧制_____１３従順な_____１４扱いやすいこと_____
１５合理的な_____１６同等_____１７補習の_____１８新兵訓練基地_____
１９群れ_____２０多選択式_____

1. abrivator 2. bloss 3. “boot camp” 4. cluster 5. constraints

6. creative 7. customers 8. dinated 9. docility 10. education

11. equivalent 12. exam 13. experienced 14. follow 15. glinder

16. gulate 17. heam 18. independent 19. instined 20. intensity

21. julique 22. leeway 23. malleable 24. minurite 25. moldarian

26. “multiple-choice” 27. nud 28. omnipresent 29. procedures 30. pindle

31. rationalized 32. recund 33. remedial 34. revictive 35. science

36. servate 37. skilled 38. sleem 39. snickle 40. spontaneity

41. strive 42. submissive 43. tailors 44. technology 45. tenure

46. “time-consuming” 47. tomby 48. tripter 49. tyranny 50. undiration

Productive Vocabulary Definition Selection Test Answer Key

１手続き29procedures ２努力する41strive　３偏在する 28omnipresent ４自発性 40spontaneity ５仕立てる 43tailors ６展性に富む23malleable ７強度 20intensity ８余裕 22leeway ９手間がかかる 46time-consuming １０制限 5constraints １１長期雇用契約 45tenure １２圧制 49tyranny １３従順な 42submissive １４扱いやすいこと 9docility １５合理的な 31rationalized １６同等 11equivalent １７補習の 33remedial １８新兵訓練基地 3bootcamp １９群れ 4cluster ２０多選択式 26multiple-choice

Appendix V Examples of Calculations for One-Sample t-Test

A. Comparison of Pilot test mean score to Population mean score
T= (Sample mean minus population mean)/sample estimate of the standard error of the mean

Sample estimate of the standard error of the mean = sample standard deviation/ the square root of the number in the sample

To test for a significant difference between the mean score of the pilot comprehension test and a population mean, it was assumed that the population mean for true random guessing would result in a mean score of 14 on a 28 item True/False test. The Sample population mean = 14.89189. Sample standard deviation = 2.30680. Number of sample 37 students.

T= (14.89189-14)/ Sample estimate of the standard error of the mean
Sample estimate of the standard error of the mean = 2.30680/$\sqrt{37}$ = .37924
T= (14.89189-14)/.37924
T= 2.35178

Degrees of Freedom = number in the sample minus one = 37-1 =36
Null hypothesis: population mean = 14.
Alternate hypothesis: population mean is not equal to 14.
Critical value of t for a two-tailed test at 30 degrees of freedom with an alpha of .05 = 2.042
Reject null hypothesis if t greater than or equal 2.042 or less than or equal –2.042
T = 2.35178 so null hypothesis is rejected; the mean score of pilot test is significantly different from the population mean.

If the apha level is set at .02, the critical value of t for a two-tailed test at 30 degrees of freedom = 2.457 in which case the mean score of the pilot test does not significantly differ from the population mean.

B. Comparison of individual items on Pilot test to individual items for Population.
The calculation and reasoning was essentially the same as above. However, to determine the sample mean for each item, the percentage of students correctly answering each individual test item was compared to a population mean of .5 which was derived by assuming that the population of a true random guessing group would on average correctly answer half the questions on a True/False test item.

Appendix VI Comprehension Test

There are two forms of the comprehension test. One contains instructions for groups with access to the text. The other contains instructions for the group not receiving the text. Questions are the same on both forms

A. Instructions for groups reading the text

Which statements are true or false about the article. For each item:
Mark Bubble [1] for a TRUE statement.
Mark Bubble [2] for a FALSE statement.
When you finish this part of the test, please raise your hand.
記事を読んで次の文が正しいか間違っているか答えなさい。
マークシートに正しい場合は（１）、間違っている場合は（２）に印をつけなさい。
テストが終わったら手を上げてください。

B. Instructions for No Text No Dictionary Group

The purpose of this test is to check your test taking skills and ability to make good guesses from background information. There are 28 statements about an article that you have not read. Which statements about the article do you think are true or false?
For each item:
Mark Bubble [1] for a TRUE statement.
Mark Bubble [2] for a FALSE statement.
Please mark **all** 28 items either true or false on the mark sheet.
When you finish this part of the test, please raise your hand.
このテストの目的はテストを受ける技術と予備知識から推測力をチェックすることです。ある記事についての文章が 28 問あります。その文章が正しいか間違っているか推測して全て答えてください。マークシートに正しい場合は（１）、間違っている場合は（２）に印をつけなさい。　テストが終わったら手を上げてください。

1. もし,この記事が新聞に載るならば社説欄にあるでしょう。
2. 記事に依ると、学生には様々なコースの選択があります。
3.シルバンラーニングセンターは、放課後に勉強する低レベルの生徒向けです。
4. 記事に依ると、教科書内容と時間割は生徒をコントロールします。
5. シルバンラーニングセンターは非営利の組織です。
6. 著者は、マクドナルドにポジティブな意見を持っているようです。
7. 記事に依ると、大学は教育をコントロールする技術を発展させました。
8. 記事に依ると、キンダーケアはたいてい幼児教育の経験豊富な人材を雇用します。
9. 記事に依ると、教授はいつクラスが終わるかをきめます。
10 著者は、おそらく記事において議論されたテクノロジーは教育においてポジティブな効果があると思っています。

11. 著者の主な点は、規則に従うことを学生に学ばせるのは学生と社会に役立つということです。
12. 著者は、授業計画に否定的な意見を持っているようです。
13. 記事に依ると、学生は暗記や客観的なテストを尊重するように教えられています。
14. 記事に依ると、比較的に未訓練の従業員は熟練者よりコントロールしやすい。
15. 記事に依ると、プロモーションシステムは教師にとって授業の準備時間をより少なくさせているかもしれない。
16. 著者は、おそらく、大学の教授と生徒の関係はファーストフードの従業員と客のように何も共通がないと思っています。
17. 記事に依ると、通常、教育システムは想像力を報います。
18. 記事に依ると、教授はコンピューターによる多選択式試験を使用しています。何故なら、学生の能力を知るのによい方法であるからです。
19. 記事に依ると、既成のカリキュラムの主な点は先生たちの時間を省けることです。
20. 記事に依ると、いくつかの保育組織はファーストフードのレストランに似ています。
21. 記事に依ると、創造力のある学生は教育のシステムにとって費用がかかります。
22. 記事に依ると、学生と教授は大学のシステムにより平等にコントロールされます。
23. 先生は授業が終わるので、カメを片付けるようにいいます。
24. 記事に依ると、教育システムは扱いやすい生徒を養成しています。
25. 記事に依ると、幼稚園と小学校では一生懸命生徒をコントロールしようとしています。
26. 記事に依ると、キンダーケアの先生は自分たちで授業内容を考えます。
27. 記事に依ると、大学生は教育のシステムコントロールに従っています。
28. 記事に依ると、全般的に、教育システムが強調されることにより、自立しない生徒が育成されています。

Answer key for Comprehension Test
1T 2F 3T 4F 5F 6F 7T 8F 9F 10F 11F 12T 13T 14T 15T 16F 17F 18F 19F 20T 21T 22F 23F 24T 25T 26F 27T 28T

Appendix VII Pre-translation Comprehension Questions with Areas of Text Targeted by Question

See appendix II for coding of text passage

Question #1 If the article were in a newspaper, it would probably be in the opinion section. True
Multiple locations in text, Inference

Question #2. According to the article, students have many choices in the types of course they can take. False
Paragraph2, Detail, No/Minimal Inference
"For example, besides the constraints7 already mentioned1, universities often give students little leeway10 in the courses they may take."
Comment: Not knowing the word leeway might result in incorrectly answering the question. This item may be more of a vocabulary question.

Question #3. Sylvan Learning Centers are for low level students that study after school. True
Paragraph 6**,** Detail, No/Minor Inference
Sylvan5 learning centers are after-school centers for remedial9 education.
Comment: Not knowing the word remedial might result in incorrectly answering the question. This item may be more of a vocabulary question.

Question #4. According to the article, textbook content and set class times exert control over students. False
Paragraph 4,Main, No/Minimal Inference
The clock and the lesson plan also exert6 control over students, especially in grade school and high school.
Comment: subsequent sentences also elaborate on this idea

Question #5. According to the article, Sylvan Learning Centers are non-profit institutions. False
Paragraph6, Main, No/Minor Inference
Through its training methods1, rules, and technologies, for-profit6 systems like the Sylvan5 Learning Center exert6 great control over their "teachers."
Comment: the word franchised in the first sentence of paragraph 6 also provides another hint.

Question #6. The author seems to have a positive opinion of McDonald's. False
Multiple locations in text, Main, Inference
Comment: McDonalds is not mentioned directly until paragraph 6 but there are references to fast-food and "McChild" in paragraph 5. The text does not explicitly state that the author has a negative opinion of McDonald's but a general overall understanding of the text would lead a reader to conclude that at best the author is neutral about McDonald's

Question #7. According to the article, universities have developed technologies to control education. True

Paragraph 1, Main, No/minimal Inference

"*Universities have developed a variety of nonhuman1 technologies to exert6 control over the process of education.*

Question #8. According to the article, Kinder-Care mostly hires people with a lot of previous childcare experience. False
Paragraph 5, Detail, No/Minimal Inference
"Kinder-care tends1 to hire1 short-term1 employees1 with little or no training in child care1."

Question #9. According to the article, professors decide when classes end. False
Paragraph 1, Detail, Minor Inference
"Class periods are set by the university and determine when classes must end. Students leave at the assigned2 time no matter where the professor happens to be in the lecture1."

Question #10 The author would probably agree that the technologies discussed in the article have a positive effect on education. False
Multiple locations in text, Main, Inference

Question #11 The author's main point seems to be that teaching students to follow rules benefits students and society. False
Multiple locations in text, Main, Inference
Comment: It may be possible to answer this based on a good understanding of some individual sentences but the idea running throughout the text is that following rules benefits institutions and not necessarily the students or customers.

Question #12. The author seems to have a negative view of lesson plans. True
Paragraph 4,Detail, Inference
"Because of the "tyranny10 of the clock," a class must last until, and end at, the sound of the bell, even though learning does not often conform3 to the clock. Thus5, even if students are just about to comprehend something, the lesson must end and the class must move on to something else. Because of the "tyranny10 of the lesson plan," a class must focus2 on what the plan requires for the day, no matter1 what the class (and perhaps the teacher) may find interesting."
Comment: Example of the turtle/crab also makes the same point.

Question #13. According to the article, students are taught to value memorization and objective testing. True
Paragraph 3, Main, No/Minimal Inference
"*Students are taught not only to obey4 authority6, but also to embrace6 the rationalized10 procedures9 of rote5 learning and objective2 testing."*
Comment: this item is concerned with an area of the text where there is a cluster of difficult words.

Question #14. According to the article, relatively untrained employees are easier to control than skilled employees. True

Paragraph 5, Main, No/Minimal Inference
Rather, relatively1 untrained2 employees1 are more easily controlled by the nonhuman1 technology of the omnipresent10 "instruction1 book."

Question #15. According to the article, the promotion system may force professors to spend less preparation time on teaching. True
Paragraph 1**,** Detail, Minor inference
"The publishing demands2 of the tenure9 and promotion1 system may force professors to devote2 far less time to their teaching than they, and their students, would like."

Question #16. The author would probably agree that professors and students at universities have nothing in common with employees and customers at fast-food restaurants. False
Multiple locations in text, Main, Inference
Comment: to answer this correctly, the student must link ideas in both the beginning and end sections of text.

Question #17. According to the article, the education system usually rewards students for creativity. False
Paragraph 3, Main, No/Minimal Inference
More important, spontaneity10 and creativity tend1 not to be rewarded6,

Question #18. According to the article, professors use computer-graded multiple choice exams because they are a good way to determine students' ability. False
Paragraph 1, Detail, Minor inference
"So, too, because the university requires grading,1 the professor must test students. In some universities, final grades must be submitted4 within forty eight-hours of the final exam, which may force professors to employ computer-graded3 multiple-choice9 exams."

Question #19. According to the article, the main benefit of a ready-made curriculum is that it saves teachers time. False
Paragraph 5, Main, Inference
What these employees1 do in the "classroom" is largely determined1 by an instruction1 book with a ready-made6 curriculum3. Staff members open the manual to find activities spelled out5 in detail1 for each day. *Clearly, a skilled, experienced, and creative teacher is not the kind of person "McChild6" care centers seek to hire1. Rather, relatively1 untrained2 employees1 are more easily controlled by the nonhuman1 technology of the omnipresent10 "instruction1 book."*
Comment: Not explicitly stated in text but should be inferable that the ready-made curriculum is for the benefit of Kindercare.

Question #20. According to the article, some child care institutions are like fast-food restaurants. True
Paragraph 5, Main, No/Minimal Inference

An even more extreme3 version of this emphasis5 appears in the child-care1 equivalent10 of the fast-food restaurant,
Comment: Depends on understanding of equivalent. Also helpful to understand what "this" refers to.

Question #21.According to the article, creative students are expensive for the education system. True
Paragraph 4, Main, No/Minimal Inference
creative, independent students are often, from the educational system's point of view1, "messy5, expensive, and time-consuming8."

Question #22. According to the article, students and professors are equally controlled by the university system. False
Paragraph 2, Main, No/Minimal inference.
"*Of course, students are even more controlled than professors by the university system."*

Question #23. The teacher tells the students to put away the turtle because it is time to end the lesson. False
Paragraph 4,Detail, No/Minimal Inference
Now children, put away the turtle, the teacher insists1. We're going to have our science lesson. The lesson is on crabs4."

Question #24. According to the article, the education system may train students to be docile. True
Paragraph3, Main, No/Minimal Inference
"and may even be discouraged3, leading to what one expert1 calls "education for docility10."
Comment: Same idea also appears in the title, "Education: learning to be docile"
Comment: this may be more of a vocabulary question because it depends on understanding the word docility.

Question #25. According to the article, kindergartens and grade schools try hard to control students. True
Paragraph3, Detail, No/Minimal Inference
Grade schools2 in particular3 have developed many technologies to control students. Many schools strive8, right from the start, to have students conform3 to their rules. Kindergarten5 has been described1 as educational "boot camp.7" Comment: this is question is drawn from illustrative example sentences but also contains the main idea of the paragraphs introductory sentence

Question #26. According to the article, Kinder-Care teachers think of their own lesson plans. False
Paragraph5, Detail, Minor Inference
What these employees1 do in the "classroom" is largely determined1 by an instruction1 book with a ready-made6 curriculum3. Staff members open the manual to find activities spelled out5 in detail1 for each day.

Clearly, a skilled, experienced, and creative teacher is not the kind of person "McChild6" care centers seek to hire1.

Comment: Not explicitly stated that teachers don't make their own lesson plans but surrounding text and vocabulary strongly suggest it

Question # 27. According to the article, university students have submitted to the education system's control. True

Paragraph 3, Main, Minor Inference

Those who conform3 to the rules are thought of as good students, while those who don't are labeled4 bad students. *Thus5, as a general rule, the students who end up6 in college are the ones who have successfully submitted4 to the control mechanisms4.*

Question #28. According to the article, overall, the matters emphasized by the education system produce students who are not independent. True

Paragraph 4, Main, Minor Inference

Overall4, the emphasis5 tends1 to be on producing submissive8, malleable10 students; creative, independent students are often, from the educational system's point of view1, "messy5, expensive, and time-consuming8."

References

Alderson, J.C. (2000) *Assessing Reading*. Cambridge: Cambridge University Press.

Albus, D., Thurlow, M., Liu, K. and Bielinski, J. (2005) 'Reading Test Performance of English-Language Learners Using an English Dictionary'. *Journal of Educational Research* 98 (4): 245-54.

Aust, R., Kelley, M. J. and Roby, W. (1993) 'The Use of Hyper-Reference and Conventional Dictionaries'. *Educational Technology, Research and Development* 41 (4): 63-73.

Bachman, L.F. (1990) *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bensoussan, M. and Laufer, B. (1984) 'Lexical Guessing in Context in EFL Reading Comprehension'. *Journal of Research in Reading* 7(1): 15-32.

Bensoussan, M., Sim, D. and Weiss, R. (1984) 'The Effect of Dictionary Usage on EFL Test Performance Compared with Student and Teacher Attitudes and Expectations'. *Reading in a Foreign Language* 2 (2): 262-276.

Elley, W. B. (1991) 'Acquiring Literacy in a Second Language: The Effect of Book-Based Programs'. *Language Learning* 41 (3): 375-411.

Gass, S. (1999) 'Discussion: Incidental Vocabulary Learning'. *Studies in Second Language Acquisition* 21 (2): 319-33.

Henriksen, B. (1999) 'Three Dimensions of Vocabulary Development'. *Studies in Second Language Acquisition* 21: 303-317.

Hill, M. and Laufer, B. (2003) 'Type of task, time-on task and electronic dictionaries in incidental vocabulary acquisition'. *International Review of Applied Linguistics in Language Teaching* 41 (2): 87-106.

Hu Hsueh-chao, M and Nation, P. (2000) 'Unknown Vocabulary Density and Reading Comprehension'. *Reading in a Foreign Language* 13 (1): 403-430. http://nflrc.hawaii.edu/rfl/html

Hulstijn, J. H. (1993) 'When Do Foreign-Language Readers Look Up the Meaning of Unfamiliar Words? The Influence of Task and Learner Variables'. *The Modern Language Journal* 77 (2): 139-47.

Knight, S. (1994) 'Dictionary Use While Reading: The Effects on Comprehension and Vocabulary Acquisition for Students of Different Verbal Abilities'. *The Modern Language Journal* 78 (3): 285-99.

Koren, S. (1997) 'Quality versus Convenience: Comparison of Modern Dictionaries from the Researcher's, Teacher's and Learner's Points of View'. *TESL-EJ* 2 (3 A2) : 1-16. http://www-writing.berkley.edu/TESL-EJ/ej07/a2.html. Accessed 2006/02/04.

Koyama, T. and Takeuchi, O. (2004a) 'Comparing Electronic Dictionaries: How the Difference Affected EFL Learning'. *JACET Bulletin* 38: 33-46.

Koyama, T. and Takeuchi, O. (2004b) 'How Look-Up Frequency Affects EFL Learning?: An Empirical Study on the Use of Handheld-Electronic Dictionaries. Pacific Association for Computer Assisted Language Learning: PacCall 2004 Online Conference Proceedings. http://www.paccall.org/past_events/2004proceedings_papers/koyama.pdf Accessed 2006/03/09.

Krashen, S. (1989) 'We Acquire Vocabulary and Spelling by Reading: Additional Evidence for the Input Hypothesis'. *The Modern Language Journal* 73 (4): 440-64.

Laufer, B. and Goldstein, Z. (2004) 'Testing Vocabulary Knowledge: Size, Strength, and Computer Adaptiveness'. *Language Learning* 54 (3): 399-436.

Luppescu, S. and Day, R. (1993) 'Reading, Dictionaries, and Vocabulary Learning'. *Language Learning* 43 (2): 263-87.

Meara, P. and Buxton, B. (1987) 'An alternative to multiple choice vocabulary tests'. *Language Testing* 4 (2): 142-154.

Meara, P. (1999) 'The Vocabulary Knowledge Framework'. unpublished paper, http://www.swan.ac.uk/cals/calsres/vlibrary/pm96d.htm Accessed 2006/03/20.

Melka, F. (1997) 'Receptive vs. productive aspects of vocabulary', in Norbert Schmitt and Michael McCarthy (eds.) *Vocabulary Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press.

Nagy, W. (1997) 'Receptive vs. productive aspects of vocabulary', in Norbert Schmitt and Michael McCarthy (eds.) *Vocabulary Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press.

Nagy, W.E., Anderson, R.C. and Herman, P. A. (1987) 'Learning Word Meanings From Context During Normal Reading'. *American Educational Research Journal* 24 (2): 237-70.

Nagy, W.E., Herman, P.A. and Anderson, R.C. (1985) 'Learning Words from Context'. *Reading Research Quarterly* 20 (2): 233-53.

Nation, I.S.P. (2001) *Learning Vocabulary in Another Language.* Cambridge: Cambridge

University Press.

Nist, S.L. and Olejnik, S. (1995) 'The Role of Context and Dictionary Definitions on Varying Levels of Word Knowledge'. *Reading Research Quarterly* 30 (2): 172-93.

Paribakht, T. S. and Wesche, M. (1999) 'Reading and "Incidental" L2 Vocabulary Acquisition: An Introspective Study of Lexical Inferencing'. *Studies in Second Language Acquisition* 21 (2): 195-224.

Read, J. (2000) *Assessing Vocabulary. Cambridge*: Cambridge University Press.

Ritzer, G. (1996) *The McDonaldization of Society: An Investigation into the Changing Character of Contemporary Social Life Revised Edition*. Thousand Oaks, California: Pine Forge Press.

Stirling, J. (2003) 'The Portable Dictionary: Faithful Friend or Faceless Foe?. *The English Language Garden* http://www.bellstudents.com/norwich/project10/portable_electronic_dictionaries.html. Accessed 2006/02/09.

Tang, G. M. (1997) 'Pocket Electronic Dictionaries for Second Language Learning: Help or Hindrance?'. *TESL Canada Journal/La Revue TESL du Canada* 15 (1): 39-57.

Weschler, R. and Pitts, C. (2000) 'An Experiment Using Electronic Dictionaries with EFL Students'. *The Internet TESL Journal* 6 (8) http://iteslj.org/Articles/Weschler-ElectroDict.html Accessed 2006/03/08.